Novel sequencing-based methods to characterise microbiomes, and in particular spore-forming bacteria, in the dairy processing chain

A thesis presented to the National University of Ireland for the degree of

Doctor of Philosophy

By

Aoife J. McHugh M.Sc

Teagasc Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland

School of Microbiology, University College Cork, Cork, Ireland

March 2020

Department head: Professor Paul O'Toole

Research supervisors: Professor Paul Cotter, Dr. Conor Feehily,

and Professor Colin Hill

Table of Contents

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study, leading to the award of PhD is entirely my own work (except as declared, hereafter) and has not been submitted for another degree, either at University College Cork or elsewhere.

Signed: 

Aoife J. McHugh

Student Number: 114223535

Date: 05/03/2020

Abstract

Spore-forming bacteria are a concern to dairy processors as a consequence of their potential to be pathogenic or to cause spoilage. Strict customer enforced limits are applied in the case of some sporeformers, but the traditional agar-based detection methods used to assess compliance have limits with respect to sensitivity and specificity. In this thesis, novel sequencing-based methods are applied to characterise microbiomes in the dairy processing chain with an emphasis on spore-forming bacteria. An initial 16S sequencing-based investigation of ultra-high temperature processing on the relative composition of thermophilic sporeformers in a dairy powder revealed that different temperatures impacted on the proportions of genera present. Shotgun metagenomic sequencing was subsequently used to identify the mesophilic sporeformers present in a specific type of dairy powder, produced monthly over one year, and investigate the functional potential of strains present to determine associated risks. A further study, tracking of the dairy microbiota from farm bulk tank milk to skimmed milk powder using 16S amplicon sequencing and shotgun metagenomic sequencing, highlighted the enrichment of spore-forming bacteria in the latter stages of dairy powder production. Finally, Oxford Nanopore's MinION sequencer was used for environmental monitoring in a dairy processing facility. Results from the long-read sequencer were comparable to those from Illumina-based sequencing and culture-based analysis. Overall this thesis highlights that sequencing-based methods could be used to provide a more in-depth understanding of the microbiota of dairy and the dairy processing environment. Such an improved understanding could allow

dairy processors to make timely, informed decisions to reduce the risk of microbial contamination.

Publications

McHugh, A. J., Feehily, C., Hill, C. and Cotter, P. D. (2017) 'Detection and Enumeration of Spore-Forming Bacteria in Powdered Dairy Products', *Front Microbiol,* 8, pp. 109.

McHugh, A. J., Feehily, C., Tobin, J. T., Fenelon, M. A., Hill, C. and Cotter, P. D. (2018) 'Mesophilic Sporeformers Identified in Whey Powder by Using Shotgun Metagenomic Sequencing', *Appl Environ Microbiol,* 84(20).

List of Abbreviations

A.M	Mesophilic spore content of August sample
A.M.P	Culturable mesophilic spore content of August sample
A.T	Thermophilic spore content of August sample
A.T.P	Culturable thermophilic spore content of August sample
ACT	Artemis comparison tool
BclA	*Bacillus* collagen-like protein A
BHI	Brain heart infusion
BLAST	Basic Local Alignment Search Tool
BR	Broad range
BT	Bulk tank
CaDPA	Calcium chelated dipicolinic acid
Ces	Cereulide
CFU	Colony forming unit
CIP	Cleaning in place
CT	Collection tanker
CT_P	Collection tankers pooled
CytK	Cytotoxin K
DAFM	Department of Agriculture, Food and the Marine
ddPCR	Droplet digital PCR
DNA	Deoxyribonucleic acid
DPA	Dipicolinic acid
dsDNA	Double-stranded DNA
EC	European Commission

ELISA	Enzyme-linked immunosorbent assay
EMA	Ethidium bromide monoazide
ENA	European Nucleotide Archive
EntFM	Enterotoxin FM
FAO	Food and Agriculture Organization
FDA	Food and Drug Administration
FSAI	Food Safety Authority of Ireland
FSANZ	Food Standards Australia New Zealand
Hbl	Hemolysin BL
HHR	High heat resistance
HS	High sensitivity
HTS	High throughput sequencing
LCA	Lowest common ancestor
LCB	Local collinear blocks
LL1	Late lactation 1
LL2	Late lactation 2
LR	Long read
MAGs	Metagenome assembled genomes
MDA	Multiple displacement amplification
MDS	Multidimensional scaling
ML	Mid lactation
MPC	Milk protein concentrate
MPI	Milk protein isolate
MYP	Mannitol egg yolk polymyxin

Nhe	Non-hemolytic enterotoxin
NPP	No pre-processing
NRPS	Non-ribosomal peptide synthase
OTU	Operational taxonomic unit
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PCSMA	Plate count skimmed milk agar
PGE	Pencil graphite electrode
PMA	Propidium monoazide
qPCR	Quantitative PCR
QRY	Query
REF	Reference
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
SASP	Small, acid-soluble spore-protein
SD	Standard deviation
SFI	Science Foundation Ireland
SM	Mesophilic enriched spore pasteurised
SMP	Skimmed milk powder
SMS	Skimmed milk silo
SRB	Sulphite reducing bacteria
SRC	Sulphite reducing Clostridia
ST	Thermophilic enriched spore pasteurised
T0	Trial 0

T1	Trial 1
T2	Trial 2
T3	Trial 3
TBC	Total bacterial count
UHT	Ultra-high temperature
USDA	US Department of Agriculture
USDEC	US Dairy export Council
WMGS	Whole metagenome sequencing
WMP	Whole milk powder
WMS	Whole milk silo
WPC	Whey protein concentrate
WPI	Whey protein isolate

List of Figures

List of Tables

Chapter 1. Detection and enumeration of spore-forming bacteria in powdered dairy products

Updated for this thesis, with the addition of section 1.5.3.3.5, since publication in Frontiers in Microbiology

(doi:10.3389/fmicb.2017.00109)

Authors: Aoife J. McHugh, Conor Feehily, Colin Hill and Paul D. Cotter

Contributions: Candidate drafted and edited the review.

Review revised and edited by CF, CH, and PDC.

1.1 Abstract

With the abolition of milk quotas in the European Union in 2015, several member states including Ireland, Luxembourg and Belgium have seen year on year bi-monthly milk deliveries to dairies increase by up to 35%. Milk production has also increased outside of Europe in the past number of years. Unsurprisingly, there has been a corresponding increased focus on the production of dried milk products for improved shelf life. These powders are used in a wide variety of products, including confectionery, infant formula, sports dietary supplements and supplements for health recovery. To ensure quality and safety standards in the dairy sector, strict controls are in place with respect to the acceptable quantity and species of microorganisms present in these products. A particular emphasis on spore-forming bacteria is necessary due to their inherent ability to survive extreme processing conditions. Traditional microbiological detection methods used in industry have limitations in terms of time, efficiency, accuracy and sensitivity. The following review will explore the common spore-forming bacterial contaminants of milk powders, will review the guidelines with respect to the acceptable limits of these microorganisms and will provide an insight into recent advances in methods for detecting these microbes. The various advantages and limitations with respect to the application of these diagnostics approaches for dairy food will be provided. It is anticipated that the optimization and application of these methods in appropriate ways can ensure that the enhanced pressures associated with increased production will not result in any lessening of safety and quality standards.

1.2 Introduction

The European Union's removal of milk quotas in April 2015 lead to a 2% increase in milk deliveries to dairies in the EU for 2015. Some countries are taking full advantage of the new limitless system in the EU, with Ireland, Luxemburg and Belgium increasing bi-monthly milk deliveries to dairies by in excess of 20% (Eurostat, 2016). Although the production rate has slowed in some other major dairy exporters, including New Zealand and Australia, the US has seen continued increases in production (DCANZ, 2016; Dairy Australia, 2015; USDA, 2016). The surplus milk produced can be processed into a wide variety of dairy products, including yoghurt, butter, cheeses and dairy powders. Dairy powders are a popular commodity due to their long shelf life, ease of storage and versatile nature. A wide variety of dairy powders can be produced, each with individual properties. These include whole milk powder (WMP), skimmed milk powder (SMP), whey protein concentrate (WPC), whey protein isolate (WPI), milk protein concentrate (MPC), milk protein isolate (MPI), casein and caseinates (Lagrange, Whitsett and Burris, 2015). Dairy powders can be used in fortification of other dairy products (Karam *et al.*, 2013), as well as an ingredient in a wide array of foods including soups and sauces, confectionary (Sharma, Jana and Chavan, 2012), infant formula, sports dietary supplements and in foods for health recovery (Lagrange, Whitsett and Burris, 2015; Gill, Rutherfurd and Cross, 2001). However, the increased production of dairy powders may create safety and economic risks to the dairy sector, specifically when controlling microbial loads in these products. Several key steps are involved in producing dairy powders including pasteurization, separation,

evaporation and spray drying (Figure 1.1). These thermal and mechanical processes can reduce the microbes present in the milk. However, spore-forming bacteria may survive. It has been shown that the spore-forming bacterial composition of raw milk differs considerably from their associated dairy powders (Miller *et al.*, 2015), highlighting that the processing of milk into powder changes the composition of the specific sporeformers present. Post-production, powders can be stored for extended periods and in the absence of water, bacterial metabolic activity and growth is limited (Deng, Li and Zhang, 2012), thus preventing spoilage and product defects. However, under these conditions, bacterial spores can remain dormant until more favorable conditions are encountered, when germination and outgrowth can proceed (Setlow, 2003; Setlow, 2014).

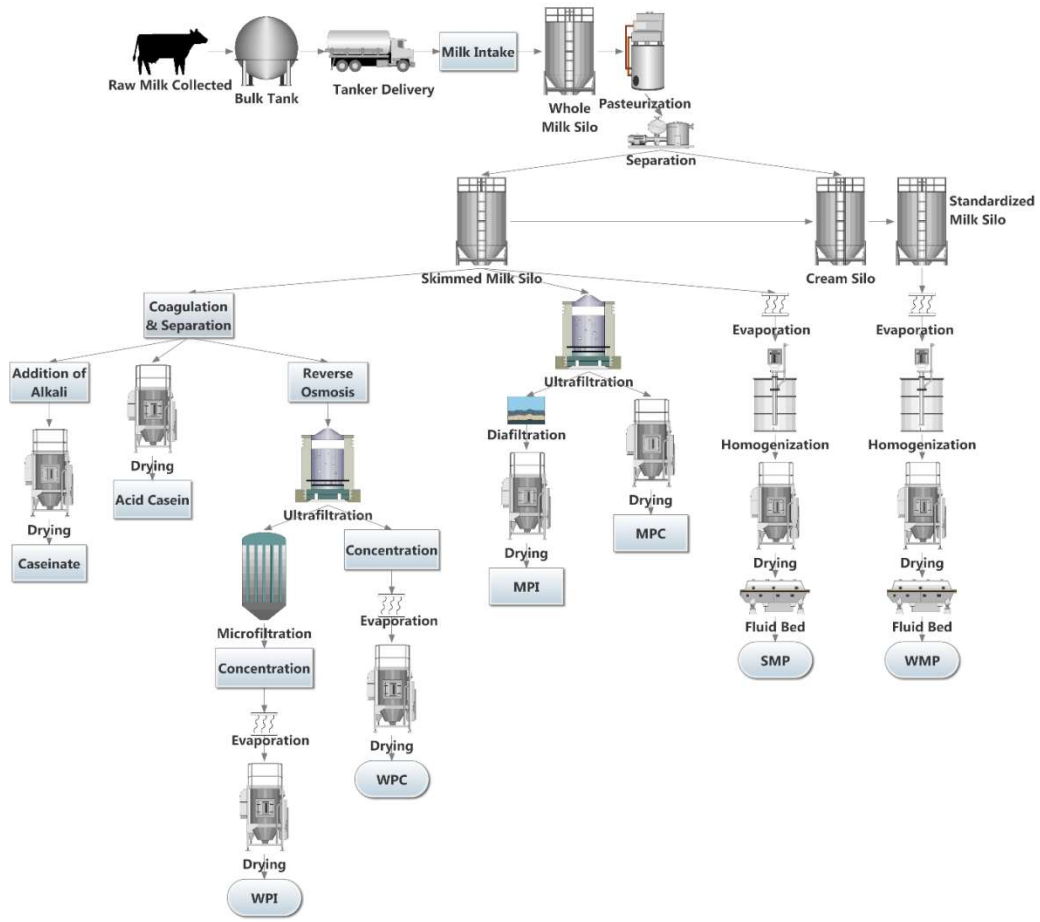


Figure 1.1 Sample dairy powder production pipelines.

1.3 Bacterial contaminants of dairy powders

1.3.1 Sources of bacterial contamination of dairy powders

Spore-forming bacteria can contaminate dairy powders through a variety of means. Bacteria can originate from the soil (Heyndrickx, 2011), faeces, bedding, feed, or milking equipment (Gleeson, O'Connell and Jordan, 2013), or can enter the raw milk *via* contaminated teats, milking cups and bulk tanks. Additionally, contamination can occur during transport from the farm to the processing plant (Pantoja, Reinemann and Ruegg, 2011), and also within the processing facility itself from poor handling and contaminated equipment (Burgess, Lindsay and Flint, 2010; Faille *et al.*, 2014). The formation of homogeneous or heterogeneous multicellular bacterial communities on the surface of processing equipment in the form of biofilms is a particular concern for the dairy processing sector and, when present, can lead to recurring problems of microbial contamination. The biofilms, which are themselves resistant to cleaning, can serve as a reservoir for bacterial spores which can slough off and contaminate dairy powders (Branda *et al.*, 2001; Faille *et al.*, 2014).

1.3.2 Common bacterial contaminants

Common contaminants identified in dairy powders include species of the class Bacilli (Table 1.1), many of which are capable of forming endospores (Checinska, Paszczynski and Burbank, 2015).

Table 1.1 Contaminants of the class Bacilli identified in powdered dairy products.

Bacilli contaminants	Reference(s)
Bacillus lichenformis	(Reginensi *et al.*, 2011; Ronimus *et al.*, 2003; Ruckert, Ronimus and Morgan, 2004; Rueckert, Ronimus and Morgan, 2005; Miller *et al.*, 2015; Sadiq *et al.*, 2016; Buehner, Anand and Djira, 2015; VanderKelen *et al.*, 2016)
Bacillus subtilis sensu lato	(Reginensi *et al.*, 2011; Ronimus *et al.*, 2003; Ruckert, Ronimus and Morgan, 2004; Rueckert, Ronimus and Morgan, 2005; Miller *et al.*, 2015; Sadiq *et al.*, 2016)
Bacillus pumilus	(Reginensi *et al.*, 2011; Ruckert, Ronimus and Morgan, 2004; Miller *et al.*, 2015; Sadiq *et al.*, 2016; Buehner, Anand and Djira, 2015; VanderKelen *et al.*, 2016)
Bacillus circulans	(Ruckert, Ronimus and Morgan, 2004; Sadiq *et al.*, 2016)
Bacillus coagulans	(Ruckert, Ronimus and Morgan, 2004; Sadiq *et al.*, 2016)
Bacillus cereus sensu lato	(Miller *et al.*, 2015; Sadiq *et al.*, 2016; Buehner, Anand and Djira, 2015; Reyes *et al.*, 2007; Zhang *et al.*, 2016)
Bacillus megaterium	(Reginensi *et al.*, 2011; Buehner, Anand and Djira, 2015)
Bacillus sonorensis	(Buehner, Anand and Djira, 2015; Sadiq *et al.*, 2016)
Bacillus altitudinis	(Buehner, Anand and Djira, 2015)
Oceanobacillus spp.	(Buehner, Anand and Djira, 2015)
Bacillus clausii	(Miller *et al.*, 2015; Sadiq *et al.*, 2016)
Bacillus thermoamylovorans	(Miller *et al.*, 2015; Sadiq *et al.*, 2016)
Anoxybacillus spp.	(Miller *et al.*, 2015; Sadiq *et al.*, 2016; Trmcic *et al.*, 2015)
Anoxybacillus flavithermus	(Reginensi *et al.*, 2011; Ronimus *et al.*, 2003; Ruckert, Ronimus and Morgan, 2004; Rueckert, Ronimus and Morgan, 2005; Sadiq *et al.*, 2016; VanderKelen *et al.*, 2016)
Geobacillus spp.	(Miller *et al.*, 2015; Trmcic *et al.*, 2015)
Geobacillus stearothermophilus	(Ronimus *et al.*, 2003; Rueckert, Ronimus and Morgan, 2005; Ruckert, Ronimus and Morgan, 2004; Sadiq *et al.*, 2016; Buehner, Anand and Djira, 2015)
Geobacillus thermoleovarans group	(Sadiq *et al.*, 2016; VanderKelen *et al.*, 2016)
Ureibacillus spp.	(Miller *et al.*, 2015)
Urebacillus thermosphaericus	(Ruckert, Ronimus and Morgan, 2004)
Aeribacillus pallidus	(Miller *et al.*, 2015; Sadiq *et al.*, 2016)
Lysinibacillus spp.	(Miller *et al.*, 2015)
Lysinibacillus sphaericus	(Sadiq *et al.*, 2016)
Paenibacillus spp.	(Miller *et al.*, 2015)
Paenibacillus cookii	(Sadiq *et al.*, 2016)
Paenibacillus macerans	(Sadiq *et al.*, 2016)
Bacillus aerophilus sensu lato	(Sadiq *et al.*, 2016)
Brevibacillus brevis	(Sadiq *et al.*, 2016)
Brevibacillus parabrevis	(Sadiq *et al.*, 2016)

Virgibacillus proomi	(Sadiq *et al.*, 2016)
Bacillus shackletonii	(Sadiq *et al.*, 2016)
Sporosarcina contaminans	(Sadiq *et al.*, 2016)
Laceyella sacchari	(Sadiq *et al.*, 2016)
Bacillus amyloliquefaciens	(VanderKelen *et al.*, 2016)

Taxa other than Bacilli have also been found to contaminate powdered dairy products with species reported including *Clostridium halophilum*, *Klebsiella oxytoca* (Buehner, Anand and Djira, 2015), *Clostridium perfringens*, *Clostridium septicum*, *Clostridium novyi/haemolyticum*, *Clostridium sporogenes* (Barash, Hsia and Arnon, 2010), *Staphylococcus aureus* (Zhang *et al.*, 2015) and *Cronobacter sakazakii* (Minami *et al.*, 2012). Bacteria of the genus *Clostridium*, as well as many of the contaminants of the class Bacilli (Table 1.1), including *Bacillus*, *Anoxybacillus*, *Geobacillus*, *Lysinibacillus*, *Brevibacillus* and *Paenibacillus*, have a considerable advantage due to being capable of forming stress-resistant endospores. These genera, and their associated species, vary considerably with respect to the range of temperatures in which they can grow, and include some psychrophilic (Ivy *et al.*, 2012) and thermophilic (Watterson *et al.*, 2014; Burgess, Lindsay and Flint, 2010) species. Dairy product contaminating sporeformers can also differ by virtue of preferring anaerobic (Doyle *et al.*, 2015) or aerobic (Gopal *et al.*, 2015) conditions. Although many sporeformers are not pathogenic and are seen primarily as indicators of poor hygiene during milk collection and or processing (Burgess, Lindsay and Flint, 2010), some can cause disease (Andersson, Rönner and Granum, 1995). Of the sporeformers identified in powders, specific representatives of *Clostridium spp*. and *Bacillus spp*. are the most worrying from a food safety point of view. *Clostridium* are anaerobic sporeformers, of which *C. botulinum* is the most notorious due to its highly potent botulinum toxin. There are many types of botulism including foodborne botulism, wound botulism, infant botulism and adult intestinal botulism. Infant botulism is the most common form (Sobel, 2005). Strains of *C. botulinum* isolated clinically have been identified in containers of opened milk

powder from the home of patients with infant botulism (Brett *et al.*, 2005; Johnson *et al.*, 2005). Despite this, and although many species of *Clostridium* have been identified in dairy powders (Buehner, Anand and Djira, 2015; Barash, Hsia and Arnon, 2010), dairy powders have never been found to be responsible for a case of infant botulism (Doyle *et al.*, 2015; Johnson *et al.*, 2005; Brett *et al.*, 2005). However, it is worth noting that anaerobic spore-forming bacteria, like *C. botulinum*, are less common than aerobic sporeformers in dairy powders. This may be due to the high degree of aeration involved in dairy powder processing or that testing criteria for sporeformers has been optimized to identify aerobic sporeformers except in the case of phenotype based assays for specific groups of anaerobic species. The ability of certain *Clostridium* species to reduce sulphite to sulphide under anaerobic conditions resulting in black colonies on specific media has been widely utilized. The accuracy of these qualitative and quantitative approaches has previously been discussed (Doyle *et al.*, 2015). Of the aerobic sporeformers identified, the majority have been of the genus *Bacillus* (Table 1.1). Many species of this genus are generally regarded as safe and some are even used as probiotics (Hong, Duc le and Cutting, 2005); e.g., Bactisubtil, Biovicerin and Biosubtyl containing *B. cereus*, Bidisubtilis containing *B. subtilis*, Biosporin and Primal Defence containing *B. subtilis* and *B. licheniformis*, Biosubtyl containing *B. pumilus*, Enterogermina containing *B. clausii* and Lactospore containing *B. coagulans* (Hong, Duc le and Cutting, 2005). Other species of *Bacillus* have been used in the production of animal feed-stuffs; e.g., *B. subtilis* has been utilized for the fermentation of indigestible by-products of soya bean oil production to yield a suitable food source for monogastric animals (Wongputtisin *et al.*, 2014). *B. cereus*

sensu lato is the most important group of species identified from a pathogenic perspective (Bottone, 2010). This group, containing up to 11 individual, highly related species (Liu *et al.*, 2015; Okstad and Kolsto, 2011), includes species that are regarded as non-pathogenic (Okstad and Kolsto, 2011). Other species include *B. thuringiensis* which is used as pesticides (Bravo *et al.*, 2013; Schnepf *et al.*, 1998); *B. cereus*, a class 2 pathogen capable of food poisoning which gave this species group its name (Bottone, 2010) and even a class 3 human pathogenic species, *B. anthracis* (Rasko *et al.*, 2005). All of these are notoriously difficult to classify and differentiate from each other (Liu *et al.*, 2015; Rasko *et al.*, 2005; Helgason *et al.*, 2000; Radnedge *et al.*, 2003). *B. cereus* is the main cause of food poisoning from within this group. *B. cereus* strains can contain many enterotoxins which are associated with diarrheal food poisoning including non-hemolytic enterotoxin (Nhe) (Lindback *et al.*, 2004; Lund and Granum, 1996), hemolysin BL (Hbl) (Beecher and Wong, 1997), and cytotoxin K (CytK) (Lund, De Buyser and Granum, 2000). It should be noted that the description of CytK as a viable enterotoxin has been called into question as, in isolation, the presence of the corresponding gene has not been linked to virulence in diarrheal pathogenesis (Castiaux *et al.*, 2015). Other molecules previously thought to be enterotoxins associated with food poisoning but which have since been reclassified include EntFM (Tran *et al.*, 2010) and BcET (Choma and Granum, 2002). Some strains of *B. cereus* also produce an emetic toxin, cereulide (Ces), a product of non-ribosomal peptide synthesis, which can cause emetic food poisoning (Toh *et al.*, 2004; Horwood, Burgess and Jane Oakey, 2004).

1.3.2.1 Spore formation

Endospores are formed in *Bacillus* and *Clostridium* species in response to environmental stress, by the activation of the master transcriptional regulator Spo0A (Hoch, 1993) following a cascade of phosphorylation including five autokinases and two phosphorelay proteins (Molle *et al.*, 2003). Spo0A binds to DNA and influences the expression of over 500 genes (Molle *et al.*, 2003). It does so directly, for example it can control efficient replication of a single chromosome for both the mother cell and fore spore by binding to the origin of replication in the mother cell (Boonstra *et al.*, 2013). But it can also work indirectly, through regulation of other transcription factors (Molle *et al.*, 2003). There are over 100 genes known to be required for spore formation, with more being identified as research in the field develops (Meeske *et al.*, 2016). Steps involved in spore formation include segregation of DNA, formation of a septum, engulfment and formation of a fore spore, formation of spore protein layers, cortex, membranes and spore coat and maturation of the spore before lysing the mother cell and being released. This process has previously been comprehensively reviewed elsewhere (Pompeo, Foulquier and Galinier, 2016; Sella, Vandenberghe and Soccol, 2014). Following its formation, an endospore can remain dormant and can persist in unfavorable environmental conditions without moisture or nutrients due to the protective structure and properties of the endospore.

1.3.2.2 Spore structure

Endospores contain several thick layers. The outer coat, or exosporium, is a thick layer only found in some species, usually those of *B. cereus sensu lato* (Lai *et al.*,

2003; Matz, Beaman and Gerhardt, 1970). The exosporium contains two layers, a basal layer surrounded by an external layer with hair like projections consisting mainly of the glycoprotein *Bacillus* collagen-like protein A (BclA) (Sylvestre, Couture-Tosi and Mock, 2002; Stewart, 2015). The exosporium, and especially BclA, contributes to hydrophobicity and aids the binding of spores to their substrates, including food preparation surfaces and stainless steel. This, along with its ability to assist spores in their avoidance of innate immune cells (Stewart, 2015), and also aids the spores' survival, spread and pathogenicity potential in the food chain. The exosporium, if present, surrounds the spore coat. The spore coat is a complex, semipermeable, proteinaceous layer found on all endospores. It is the outermost layer of *B. subtilis* spores (Setlow, 2006) and gives resistance to chemicals and enzymes, as well as structurally holding the spore together. It excludes large molecules, while allowing nutrients pass through and interact with germination receptors deeper in the spore structure (Lai *et al.*, 2003; Driks, 2002). The spore coat surrounds an outer membrane, which encapsulates the cortex. The cortex is made of specific peptidoglycan (Popham, 2002) that is assembled into rod shaped structures, located perpendicularly to the spore surface (Li, Burggraf and Xing, 2016). It confers resistance to wet heat and is essential in the dormancy of the spore as well as reducing the water content of the core (Setlow, 2006). The cortex surrounds the germ cell wall, which becomes the bacterial cell wall following germination (Setlow, 2006; Wells-Bennik *et al.*, 2016). The germ cell wall surrounds an inner membrane. This too protects the bacterial spore against chemicals, and contains the proteins required for germination back to active cells (Setlow, 2003). Proteins include transporters (some of which are associated with efflux processes

and unique to the spore inner membrane), proteases (essential for sporulation and germination), DNA repair and replication enzymes (including nucleotide excision repair enzymes, spore specific lyases and endonucleases), heat shock proteins and proteins involved in control of cellular processes in response to stress (including, but not limited to UV and oxidative stress) have all been identified in the spore inner membrane (Zheng *et al.*, 2016). These all contribute to the resistance and persistence of spores in unfavorable conditions. Inside the inner membrane is the core of the endospore, which is severely dehydrated and compacted. This dehydration allows immobilization of proteins, preventing their coagulation following heat denaturation (Sunde *et al.*, 2009). The core also contains high levels (up to 15-25% of the spores dry weight) of dipicolinic acid (DPA), most of which is chelated by divalent ions, allowing protection of spore DNA from external stressors as well as synthesis of new DNA in response to UV radiation (Setlow, 2006; Sunde *et al.*, 2009; Setlow, 2007). Also found in the spore core of *Bacillus* species is a group of small, acid-soluble spore proteins (SASP) of the α/β-type. These bind DNA in the spore core and alter its structure, thus aiding its resistance to heat, chemicals, UV radiation and osmotic pressure (Setlow, 2006; Setlow, 2007).

1.3.2.3 Survival of spore-forming bacteria in processing environments

Spores can survive processing to which vegetative cells would normally succumb. Such processing-related stressed include desiccation, dry and wet heat, UV radiation, mechanical agitation, γ-radiation, chemical exposure and hydrostatic and osmotic pressure (Nicholson *et al.*, 2000; Setlow, 2006). Indeed, while the temperatures and drying conditions used in the processing of milk to powders kills

most vegetative bacterial cells, it also inadvertently selects for these sporeformers. Once powders are rehydrated, the spores may germinate by activation of germination receptors, either in response to nutrients called germinants (Setlow, 2003) or by heat activation (Luu *et al.*, 2015). Germination independent of these receptors may also be triggered by calcium chelated dipicolinic acid (CaDPA), dodecylamine, or peptidoglycan fragments, although these mechanisms may not be applicable to the food industry (Setlow, 2014). Germination initiated by high pressure, either by activation of germination receptors or independent of them, can also occur (Setlow, 2014). Following germination, these sporeformers can proliferate in the absence of competition from other bacteria that were eradicated during processing (Brown, 2000).

1.4 Legislation governing bacterial contamination in dairy powders

Guidelines governing the levels and types of bacteria permitted in dairy powders are not very comprehensive, except in the case of infant formula. There are many different governing bodies that have set testing parameters; including the U.S. Food and Drug Administration (FDA), Food Standards Australia New Zealand (FSANZ) and The European Commission (EC). In Ireland, the Food Safety Authority of Ireland (FSAI) implements limits based on the Commission Regulation (EC) No 2073/2005(European Commission, 2005). FSAI state that aerobic colony counts in dairy powders should ideally be $< 10^4$ CFU g^{-1} (FSAI, 2014). However, this is not a legal obligation, and does not mean that the food is unsafe as characterization of the species isolated would need to be performed in order to determine product safety. The U.S. Department of Agriculture (USDA) implements the following

microbial limits in US extra grade dairy powders using the standard plate count; dry buttermilk < 20,000 CFU g^{-1} (USDA, 2001a), dry whey < 30,000 CFU g^{-1} (USDA, 2000), dry whole milk < 10,000 CFU g^{-1} (USDA, 2001b), dry casein (acid) < 30,000 CFU g^{-1} (USDA, 1968), instant non-fat dry milk < 10,000 CFU g^{-1} (USDA, 2013), non-fat dry milk (roller dried) < 50,000 CFU g^{-1} (USDA, 1984) and non-fat dry milk (spray process) < 10,000 CFU g^{-1} (USDA, 2001c). The US Dairy Export Council (USDEC) implements limits for US dairy powders destined for international customers with limits on aerobic sporeformers set to between < 500 CFU g^{-1} and < 1000 CFU g^{-1} for thermophilic and mesophilic spores, respectively, in skimmed milk powder, non-fat dry milk and whole milk powder destined for infant powder, and < 500 CFU g^{-1} and < 2000 CFU g^{-1}, respectively, in skimmed milk powder and whole milk powder (Watterson *et al.*, 2014).

In Australia and New Zealand, state agencies enforce limits set by FSANZ. *B. cereus* must be < 100 CFU g^{-1} in 4/5 samples, and < 1,000 CFU g^{-1} in 1/5 samples in dried milk powder and powdered infant formula products with added lactic acid producing cultures, and must be absent in 5 samples of 1 g in powdered infant formula. The European Commission regulation, as amended (European Commission, 2005) sets similar legal microbiological criteria including a limit of < 50 CFU g^{-1} presumptive *B. cereus* in 4/5 samples and < 500 CFU g^{-1} in 1/5 analyzed is set in accordance to EN/ISO 7932 (Standards, 2004).

Due to the competitive market for dairy ingredients, individual purchasers often set their own microbiological limits to ensure high standards. In many cases dairy powders will not receive any further treatments before incorporation into other

products. For example, powdered infant formula manufacturers often have close relationships with the dairy powder supplier to ensure high microbiological standard are met, and set strict criteria (Kent *et al.*, 2015).

1.5 Detection of spore-forming bacteria

Apart from dairy powder that is due for export from the US, no legislation thoroughly covers the enumeration or identification of all sporeformers in dairy powder. This is in spite of recent research highlighting the need for accurate spore quantification and identification (Burgess, Lindsay and Flint, 2010). Identification and enumeration of all sporeformers present in dairy powders allows identification of potential problematic species whether from a hygiene, quality or pathogenic perspective. This information would allow manufacturers implement more comprehensive and/or directed preventative measures (Pennacchia, Breeuwer and Meyer, 2014) resulting in continued economic and safety confidence in the sector. Understanding composition of total sporeformers within a product contributes to a clearer understanding of the source of potential quality or safety issues should they arise and allows faster implementation of control measures (Burgess *et al.*, 2010, Pennacchia *et al.*, 2014). Indeed, efforts have continued to be made in recent years to improve the detection and identification of spore-forming bacteria present in dairy powders (Watterson *et al.*, 2014; Sadiq *et al.*, 2016; Miller *et al.*, 2015).

1.5.1 Culture based methods

1.5.1.1 Spore count methods

Typical spore count tests involve the heating of a reconstituted powder sample to 80°C for 12 min before cooling, culturing and enumerating colonies (Watterson *et al.*, 2014; Frank and Yousef, 2004). Highly thermo-resistant spores are selected by heating to 100°C for 30 min before cooling and culturing while numbers of especially thermo-resistant spores are quantified by heating to 106°C for 30 min, cooling and culturing. Media is incubated in the presence or absence of oxygen to select for aerobic or anaerobic spore-forming species, respectively. Incubation can also be at different temperatures. Incubation at 6°C will select for psychrophilic sporeformers, incubation at 30-35°C will select for mesophilic sporeformers and incubation at 55°C will select for thermophilic sporeformers (Watterson *et al.*, 2014; Kent *et al.*, 2016). Further analysis of isolated colonies is required in order to determine the species present, and the options available for this analysis are discussed at a later stage in this review (Section 1.5.1.2, 1.5.3.1). Total bacterial counts and spore counts, although informative, are not without their limitations. Almost a century ago it was highlighted that different media will result in different bacterial counts (Ayers and Mudge, 1920) and that more than just quantitative data is needed with respect to contamination of dairy products, in order to determine the significance of the contamination (Ayers and Mudge, 1920). The use of various heating methods is somewhat redundant in terms of identification of different species (Miller *et al.*, 2015). However the actual abundance of these spore-forming bacteria does differ depending on the test method used (Kent *et al.*, 2016). In order

to get a clear picture of the total sporeformer composition present in a powder sample through culture-based approaches, a variety of incubation conditions, temperatures, agars and, possibly, heat treatments would be needed. This highlights the need for stronger/more robust test methods to determine the abundance of (spore-forming) bacteria in dairy powders.

1.5.1.2 Culture-based identification of spore-forming species

Numerous culture-based tests have been developed in order to help identify spore-forming bacteria. These involve the use of selective media and, in some cases, additional tests to provide further information regarding the identity of the species present. Both Bacara and Mannitol Egg Yolk Polymyxin (MYP) agars have been developed for the isolation of *B. cereus.* The testing used for presumptive *B. cereus* in Europe (Standards, 2004) involves the use of MYP agar and the hemolysis test. However, MYP has been shown to be not as selective as Bacara agar for *B. cereus* (Tallent *et al.*, 2012), potentially leading to false positives. Some *Clostridium* species, the sulphite reducing Clostridia (SRCs), have the ability to reduce sulphite to sulphide under anaerobic conditions. A number of sulphite containing agars have been developed for their selection (Gibbs and Freame, 1965; Weenk *et al.*, 1995; Wilson and Blair, 1924). SRCs are identified by a black color change, however other bacteria capable of reducing sulphite and can also grow on these media, these are referred to as sulphite reducing bacteria (SRBs) (Doyle *et al.*, 2015). Other tests can involve analyzing phenotypes by visualizing morphological properties and performing biochemical tests to narrow down the possible species (Reyes *et al.*, 2007; Janda and Abbott, 2002).

1.5.1.3 Limitations of culture-dependent analysis

A common limitation with all of the aforementioned methods is a requirement that the bacteria first be cultured. This can result in important difficult-to-culture species being overlooked due to inappropriate culturing conditions, temperature, aeration, and/or media type. Furthermore, colony selection may favor the selection of the largest/most plentiful colonies above the smaller/less plentiful types. Although these methods allow isolation and enumeration of culturable species, accurate identification of each species present is difficult, very time-consuming, labor intensive and can be biased. The aforementioned isolation methods can be coupled with the following, more recently developed, protein- and DNA-based methods, to provide more robust identification.

1.5.2 Protein-based methods

1.5.2.1 Enzyme immunoassays

A sandwich Enzyme-Linked ImmunoSorbent Assay (ELISA) has been developed for the detection of whole cells of *B. cereus*, by recognizing surface antigens specifically associated with *B. cereus* cells. This assay was developed by multiple location immunization of animal models with whole cell immunogen to develop hybridomas and subtractive screen was used to eliminate cross reactivity with closely related species (Zhu *et al.*, 2016b). The subtractive screen ensured the mAbs are highly specific against *B. cereus* and the assay has a lower detection limit of 0.9×10^3 cells ml^{-1} in phosphate buffered saline. This assay has been tested using food samples spiked with various pathogens without the need for culturing. It was highly effective at identifying *B. cereus* cells in mixed samples, without interference by the

food matrix or influence by other related species. Although this ELISA for detection of surface antigens is specific for *B. cereus*, it is not clear if it can recognize spores as well as vegetative bacteria, or if it can distinguish between live and dead *B. cereus* (Zhu *et al.*, 2016b). Failure to detect spores could lead to a false negative result, whereas detection of free floating antigens from dead *B. cereus* cells could lead to false positive results. Additional culturing may be needed to detect cell numbers below the lower detection limit, and thus eliminate these concerns. Enzyme immunoassays have also been developed for the detection of *B. cereus* toxins (Wehrle *et al.*, 2009; Cui *et al.*, 2016). Specific conditions are needed to ensure efficient protein production. Casein hydrolysate-glucose-yeast with 1% glucose is used for the production of enterotoxins in *B. cereus*, and 10% skim milk medium is used for cereulide production in *B. cereus* (Cui *et al.*, 2016). A negative result from a proteomic based assay would not imply that the bacteria is not present, rather protein synthesis might not be currently active.

1.5.2.2 Limitations of protein based methods

The requirement for correct expression conditions in order to identify proteins of interest is a hugely limiting step in protein based method for species identification. This is particularly true for spore-forming bacteria, whose presence is of concern but are currently in a dormant state during sample testing. Such requirements for specific growth conditions increase the analysis time and complexity, which may not be possible for large scale analysis of many possible toxin producers in laboratory situations. Furthermore, it is expected that the proteinaceous nature of

dairy samples would greatly impeded the sensitivity of any protein analysis performed without initial culturing, even if expression was occurring.

1.5.3 DNA-based methods

1.5.3.1 Post culture DNA-based classification methods

1.5.3.1.1 Random amplified polymorphic DNA polymerase chain reaction (RAPD-PCR)

RAPD-PCR uses short random primers to amplify multiple random DNA segments which, once visualized on an agarose gel, give unique patterns (Williams *et al.*, 1990). Analysis of these fingerprints allows differentiation of species and strains by comparing profiles of various known strains. (Ronimus, Parker and Morgan, 1997; Ronimus *et al.*, 2003). This method has been applied to colonies obtained from dairy powders in New Zealand to identify *Geobacillus stearothermophilus*, *Anoxybacillus flavithermus*, *Bacillus licheniformis*, and *Bacillus subtilis* as the main contaminants of whole milk powders and skimmed milk powders, as well as buttermilk and goat milk powders (Ronimus *et al.*, 2003). It has also been applied to whole and skimmed milk powders in Uruguay, correctly identifying the presence of *B. licheniformis*, *B. megaterium*, *B. pumilus*, *A. flavithermus* and *B. subtilis* (Reginensi *et al.*, 2011). Indeed, using this approach, *G. stearothermophilus*, *A. flavithermus* and *B. licheniformis* have been identified as the dominant species in whole and skimmed milk powders from multiple countries including; Poland, Germany, Switzerland, France, Portugal, Netherlands, Great Britain, Ireland, Canada, USA, Mexico, Chile, Brazil, South Africa, Thailand, Australia and New Zealand. *B. subtilis*, *Bacillus circulans*, *Ureibacillus thermosphaericus*, *Bacillus*

coagulans and *Bacillus pumilus* have also being identified, albeit in lower quantities (Ruckert, Ronimus and Morgan, 2004). A common feature of the RAPD-PCR approach is the highlighting of the 3-4 most dominant species. However, species of lower abundance might be the most interesting in terms of food security and spoilage. One study described the use of RAPD-PCR, and revealed a more in depth array of species, in Chinese dairy powders (Table 1.1) (Sadiq *et al.*, 2016). Apart from identifying previously unreported species, other details worth noting are that *B. licheniformis*, *G. stearothermophilus*, and *A. flavithermus* were again established as being present in high abundance while, importantly, *B. cereus* group species were also identified. This observation obviously has implications for food safety (Sadiq *et al.*, 2016). Although informative, analysis of the gel bands in RAPD PCR is very subjective allowing errors in classification and bias. Furthermore, the method requires time-consuming and laborious preparation of reference strains and there may also be variability between gels with the same samples, thus large-scale analysis would be difficult.

1.5.3.1.2 Sequencing housekeeping genes

Housekeeping genes are genes that are essential for the functions of the cell and viability of the organism, and thus typically contain highly conserved regions (Gil *et al.*, 2004; Eisenberg and Levanon, 2013). Genes that contain such highly conserved regions at either end of a more variable region are particularly useful for strain identification purposes as the conserved regions can be targeted using degenerate primers to facilitate PCR amplification and sequencing of the variable region (Case *et al.*, 2007). Identification of genera present is facilitated by comparison with

databases of corresponding variable region sequences of known origin (Case *et al.*, 2007). Many genes have been utilized for classification of species in fluid milk in the form of molecular typing (Durak *et al.*, 2006). Other typing methods have been described for milk powder isolates of *Geobacillus spp.* and *B*. *licheniformis* based on variable number tandem repeat analysis (Seale *et al.*, 2012; Dhakal *et al.*, 2013). The 16S rRNA gene is ubiquitous among bacteria, and contains multiple conserved and variable regions making it extremely useful, in general, for taxonomic classification. However, 16S rRNA gene sequencing cannot differentiate between closely related species or subtypes and other housekeeping genes such as *gyrB* or *rpoB* have been utilized to do so (Case *et al.*, 2007; Durak *et al.*, 2006). Recently, both the *rpoB* and 16S rRNA genes have been used to characterize the contaminating psychrophilic, mesophilic and thermophilic spore populations isolated from sweet whey, WPC, non-fat dry milk and acid whey powders. At least 14 different species were identified, with *B. licheniformis*, *Geobacillus spp*. and *Anoxybacillus spp*. being the most abundant (Miller *et al.*, 2015). These methods have the potential to allow identification and monitoring of persistent species and subtypes throughout dairy powder processing plants (Seale *et al.*, 2012; Dhakal *et al.*, 2013). Although not currently employed in sequencing dairy powder isolates, *cpn60* (Schellenberg *et al.*, 2011; Durak *et al.*, 2006), *pycA*, *ccpA* (Liu *et al.*, 2015) and *groEL* (Chang *et al.*, 2003) have all been used to varying success in the sequencing of isolates from fluid milk (Durak *et al.*, 2006), vaginal (Schellenberg *et al.*, 2016) and marine (Liu *et al.*, 2013) populations and remain as potential targets for future application to study dairy powder-associated microbes.

1.5.3.1.3 Pyroprinting

Pyroprinting utilizes sequencing by synthesis on multiple copy polymorphic loci simultaneously. The sequence reads are digitalized and can be compared using Pearsons correlation distance matrix to identify strains (Black *et al.*, 2014). This method has been developed and utilized for source tracking i.e. tracing sources of microbial contamination in end products or, more specifically, of endospore-forming Bacilli in raw milk through to dairy powders. Presumptive species identified in powder included *Geobacillus thermoleovorans*, *A. flavithermus*, *B. licheniformis*, *B. pumilis* and *B. amyloliquefaciens* (VanderKelen *et al*., 2016). These results correlate well with previous studies on raw milk and powders using the Sanger sequencing approach (Durak *et al.*, 2006; Miller *et al.*, 2015).

1.5.3.1.4 Limitations

All of the above tests allow identification of the most abundant culturable species identified in dairy powders. However, they are limited by an initial requirement for culturing and, unless these methods are modified for identification of species directly from dairy powders, they are not suitable for the identification of non-culturable species or species of lower abundance which can be out competed when culturing, unless selective media is employed. Ultimately, while promising, these methods when compared to culture-independent sequencing (Section 1.5.3.3.1) are labor intensive and time consuming.

1.5.3.2 Targeted DNA based approach

A more targeted approach can be taken in the food sector to detect specific pathogens or groups of interest. These assays allow detection of toxin genes, possible pathogenic groups or members of a species of interest. Most of these have been adapted to allow amplification directly from mixed DNA extracted from foodstuffs and thus avoid the limiting step of culturing. Many also allow quantification of the species/toxin gene containing group. Of particular relevance to this review is the fact that a great deal of research has been performed with respect to such assays and the *B. cereus sensu lato*.

1.5.3.2.1 PCR assays

Polymerase Chain Reaction (PCR)-based assays have been developed for the detection of *B. cereus* toxin genes. Taqman quantitative PCR (qPCR) assay of a single component of the hemolysin toxin gene in *B. cereus* has been developed (Cattani *et al.*, 2016), amplifying the sequence corresponding to one component of one tripartite toxin. It has been reported that the Taqman probe is specific for *B. cereus* strains that contain this gene, however, not all *B. cereus* strains contain the hemolysin gene (Cui *et al.*, 2016). This assay reportedly does not give false positives with related species, such as other members of the *B. cereus sensu lato* including *B. thuringiensis* and *B. mycoides*. However this assay could lead to false negatives. The assay may fail to detect other species that have the toxin genes, or other strains of *B. cereus* that do not have this particular toxin, but may be pathogenic due to the presence of other toxins. This assay also gives accurate quantification of viable *B. cereus* by comparison to standard curves. Multiplex endpoint PCR of toxin genes

has also been performed to identify *B. cereus* in dairy samples. These assays included primers to amplify single components of *B. cereus* enterotoxin genes, i.e. those encoding Nhe, CytK and Hbl (Zhang *et al.*, 2016) as well as enterotoxin FM (EntFM) and emetic toxin Ces (Forghani *et al.*, 2015). However, the specificity of these assays was only tested using *B. cereus* and non-*Bacillus* species. Multiplex PCR of multiple components of *B. cereus* toxin genes has also been performed on single bacterial colonies isolated from dairy products and environments (Wehrle *et al.*, 2009). This approach allows detection of all components needed to produce viable enterotoxins, and thus lessening the chance of false readings compared to other assays that only identify one toxin gene component. Multiplex endpoint PCR assays have also been developed for hygiene indicator species, *G*. *stearothermophilus* and *A*. *flavithermus* isolated from dairy powders. These assays rely in the species specific conserved regions of ITS 16S-23S rRNA region and the rpoB gene (Pennacchia, Breeuwer and Meyer, 2014). Further validation of these assays could lead to their use on DNA isolated directly from dairy powders. Finally, droplet digital PCR (ddPCR) allows precise, absolute quantification of a target DNA sequence. The DNA is encapsulated into many water in oil emulsion droplets and a PCR performed on each (Pinheiro *et al.*, 2012). This culture-independent method has recently been used to detect *B. cereus* in fluid milk and can provide absolute quantification without need for comparison to standard curves. In this instance ddPCR was implemented using primers that target the *gyrB* gene of *B. cereus sensu lato* and the assay was found to have a lower detection limit than traditional qPCR (Porcellato, Narvhus and Skeie, 2016), which is ideal for dairy powders that have low levels of contamination.

1.5.3.2.2 Biosensors

The assays described above also have the potential to be employed in the form of biosensors. Indeed, biosensors are already being developed for detection of a toxin gene found in *B. cereus* in milk and powder (Izadi *et al.*, 2016). These biosensors are DNA based pencil graphite electrode (PGE) biosensors, in which a *nhe* toxin gene primer is immobilized on gold nanoparticles. Positive results are measured by an increase in charge resistance on the biosensor from the hybridization of the target DNA to *nhe* toxin sequence.

1.5.3.2.3 Limitations of targeted DNA assays

Although these methods do not give a complete view of the microbial composition in a dairy powder, they are useful as a test for key spoilage and pathogenic bacteria, including producers of harmful toxins. It is important to note that *B. cereus sensu lato* toxin genes are not specific to any one species of the group, nor is one toxin found in all *B. cereus* (Liu *et al.*, 2015; Cui *et al.*, 2016; Zhu *et al.*, 2016a). However, targeting toxins allows detection of all possible pathogenic species. Singleplex assays that target one component of one toxin may be prone to false negatives (Cui *et al.*, 2016), i.e. producers of other toxin types being overlooked, thus underestimating the number of pathogenic *B. cereus* cells in a sample. Multiplex assays targeting many toxins, are more robust and can be beneficial for the food industry as they are a good indicator of potential food pathogens. Targeting all components of a toxin system may be required to confirm if there is a true potential for toxin production. Furthermore, while the genes for toxins may be present, it is unclear from these assays whether any active proteins are functionally

expressed. The alternative use of a non-toxin gene for identification of *B. cereus* (*gyrB*) does not distinguish between members of *B. cereus sensu lato*, nor does it identify if the species identified are capable of being pathogenic. Overall the detection of toxin and species specific genes are a good indicator of potential pathogenic and other species of interest being present. Although issues remain, future improvement and development should result in the full potential of these approaches being realized.

1.5.3.3 Culture-independent, non-targeted DNA analysis

As outlined, there are limitations associated with the aforementioned culture-dependent and targeted assays. Culture-independent DNA-based analysis should be considered when striving to obtain an overview of all (i.e. culturable and non-culturable) spore-forming species present in dairy powders. This involves a shift away from testing for and identifying only specific known spore-forming bacteria in order to eliminate the possibility of currently unknown or underappreciated microbiology-related food security threats.

1.5.3.3.1 Next generation sequencing for the identification of dairy powder contaminants.

In the last decade, considerable advances have meant that next generation DNA sequencing platforms have surpassed traditional Sanger sequencing platforms in terms of speed and potential applications. Their initially extremely short sequencing read lengths are less of a concern as sequencing lengths of Illumina and Ion platforms have increased (Quail *et al.*, 2012) and new, even longer read, platforms have been developed by PacBio and Oxford Nanopore (Madoui *et al.*, 2015; Quail *et*

al., 2012). The advantages and disadvantages of the various sequencing platforms have been previously reviewed elsewhere (Goodwin, McPherson and McCombie, 2016). Regardless, research laboratories now have a much greater choice when determining which sequencing technology to use, though it should be noted that results generated using different methods, technologies or bioinformatics pipelines are not always consistent (Clooney *et al.*, 2016). Whole genome shotgun sequencing is the process by where the whole genome of a single colony is sequenced. The DNA is extracted and sheared it into small pieces, before sequencing of these pieces and the use of computer software to assemble these sequences reads back together. This process can be applied to metagenomes, the term used to denote all of the genomic information from an entire community of different cells, for example the contaminants in dairy powders (Sharpton, 2014). The application of metagenomic techniques to the analysis of dairy products presents exciting opportunities. Metagenomic sequencing eliminates the need to culture, thus reducing bias, and allows the identification of species that are difficult to, or cannot be, cultured in the laboratory. Metagenomic sequencing has been applied to single gene products, such as the aforementioned 16S rRNA gene that can differentiate between all bacteria present to the genus level, while the *spo0A* gene has been targeted to specifically identify spore-forming Firmicutes in mixed populations. A whole metagenome 'shotgun', i.e. untargeted, approach has also been attempted and comparison of 16S amplicon sequencing, *spo0A* amplicon sequencing and metagenomic shotgun sequencing performed for the identification of Firmicutes in metagenomic samples (Filippidou *et al.*, 2015). Each method has advantages and disadvantages. Amplicon sequencing is more cost effective, high

throughput and rapid but often only gives accurate classification to genus level, and may over-estimate microbial diversity in the sample (Poretsky *et al.*, 2014; Acinas *et al.*, 2004). In contrast, shotgun sequencing is more expensive, less samples can be analyzed at one time, but it gives the opportunity to accurately classify to species level provided there are accurate reference databases to compare sequence reads to (Sharpton, 2014). Shotgun sequencing also reduces the bias of amplicon sequencing that can arise due to need for an initial PCR amplification and, where relevant, variable gene copy numbers (Sharpton, 2014; Brooks *et al.*, 2015). The other advantage of shotgun metagenomic approaches is that additional information regarding other genes of interest within the microbial community can be generated. Such genes include toxin genes (Leonard *et al.*, 2015; Steffen *et al.*, 2012), sporulation genes (Filippidou *et al.*, 2015), non-ribosomal peptide synthase (NRPS) gene clusters (Schirmer *et al.*, 2005), antibiotic resistance genes (Bengtsson-Palme *et al.*, 2014), and phage genes (Dutilh *et al.*, 2014), all of which may be interesting from a food safety point of view. The sequencing reads from this approach can be difficult to analyze as they can be biased towards genomes of higher abundance. This is a particular issue when studying samples from specific human and animal microbiomes where there is a considerable amount of DNA from host cells present (Feehery *et al.*, 2013). It is important to note that, due to the high sensitivity of shotgun metagenomic sequencing, care needs to be taken to ensure the absence of contaminating cells or DNA from other environments (Salter *et al.*, 2014; Glassing *et al.*, 2016).

Regardless of the sequencing approach taken, bioinformatic expertise is needed to analyze sequencing data and compare sequence reads to databases. Databases and bioinformatics software are updated continuously and newer, more accessible programs are constantly being developed (Vincent and Charette, 2015), including more targeted programs and databases specifically for food microbes (Vangay *et al.*, 2013; Parente *et al.*, 2016).

1.5.3.3.2 Limitations of next generation sequencing

Both amplicon and shotgun metagenomic sequencing reveal the relative abundance of bacteria in a sample. Furthermore, the quantification of total bacterial load can be achieved by coupling these techniques with qPCR or ddPCR analysis (Porcellato, Narvhus and Skeie, 2016).

While the benefits of next generation sequencing in determining the safety and quality of dairy powders provide cause for optimism, there are several hurdles. Culture-independent DNA analyses rely on one's ability to extract all genomic DNA directly from the substrate for analysis. Extracting DNA from dairy powder can be difficult, especially from spore-forming bacteria. Although, many studies have endeavored to optimize methods for the extraction of DNA from spores that have been spiked into food, success has been varied (Mertens *et al.*, 2014; Wielinga *et al.*, 2011). Furthermore, the bacterial load is likely to be lower in dried dairy powders than other environmental samples in which this sort of analysis has been previously performed, such as the gut (Gill *et al.*, 2006), soil (Fierer *et al.*, 2012) and fermented food (Jung *et al.*, 2011). Low DNA concentration can be overcome through use of whole genome amplification kits (Yokouchi *et al.*, 2006; Binga,

Lasken and Neufeld, 2008). Although expensive, these provide for culture independent non-targeted analysis of all bacteria present in dairy powders even if present at low cell numbers. However, these kits are notoriously susceptible to contamination (de Bourcy *et al.*, 2014) and, ideally, ultra clean laboratory environments are needed for their use (Weinmaier *et al.*, 2015).

1.5.3.3.3 Isolation of DNA solely from sporeformers

There may be instances where there is a specific desire to specifically focus on the sequencing of DNA from the spore-forming community within a powder sample. Isolation of DNA solely from spores/spore-forming bacteria is a challenge. One possible method would be to perform standard spore pasteurization at 80°C for 12 min as described in section 1.5.1.1 above (Watterson *et al.*, 2014; Frank and Yousef, 2004) or other forms of targeted vegetative cell lysis (Wunderlin *et al.*, 2016). However, free DNA could still be present in the samples from the lysed vegetative cells. Elimination of this signal could be performed using an intercalating dye (described below). Post heat treatment, subsequent culture-based enrichment could be employed prior to DNA extraction (Watterson *et al.*, 2014; Frank and Yousef, 2004) but, as described with respect to the culture-based approaches, this has the potential to lead to bias.

Sequencing-based approaches can also be adapted to specifically focus on sporeformers by, for example targeting of the *spo0A* gene for amplicon sequencing, or through focusing specifically on this gene from within shotgun sequence data. However, yet again, the need to ensure optimal DNA extraction and the removal of DNA from dead cells is a key consideration. A less conventional way of overcoming

such challenges could involve the isolation of spores from dairy powder using density gradient centrifugation (Tamir and Gilvarg, 1966).

As noted above, free DNA from lysed vegetative cells can be present in samples following heat-treatments. Elimination of this signal could be performed using an intercalating dye. The use of intercalating dyes is especially relevant in the case of amplicon metagenomic sequencing where PCR amplification is performed (Rudi *et al.*, 2005). This has been performed utilizing the dyes propidium monoazide (PMA) or ethidium monoazide bromide (EMA) to bind free DNA in the samples (Zhang *et al.*, 2016; Forghani *et al.*, 2015; Cattani *et al.*, 2016; Rudi *et al.*, 2005). Further testing and optimization would be needed to determine if its results are as promising for dairy powder samples with mixed populations. There are contradicting studies with regard to whether EMA or PMA is best for particular applications (Wu, Chen and Levin, 2015; Seinige *et al.*, 2014). Very few studies have compared EMA and PMA in mixed populations, though EMA was reported to be favorable at penetrating heat damaged bacterial cells in fish fillets (Lee and Levin, 2009). EMA has been known to penetrate some live bacteria (Seinige *et al.*, 2014; Nocker, Cheung and Camper, 2006) whereas PMA has been seen not to penetrate all dead cells (Cattani *et al.*, 2016). The concentrations of EMA used has seen a decrease in recent years (possibly to circumvent the penetration of live cells) and, so, while early studies used 100 μg ml^{-1} (Nocker and Camper, 2006; Rudi *et al.*, 2005), more recent studies used 8-10 μg ml^{-1} (Seinige *et al.*, 2014; Wu, Chen and Levin, 2015). Alternatives, including the use of platinum (Soejima *et al.*, 2016) to bind extracellular DNA, appear promising as they have been reported to be more

selective at differentiating live/dead *E. coli* and *C. sakazakii* than PMA in water and milk. Ultimately, optimization needs to take place to develop the system that is best suited to the low microbial load of mixed populations present in powdered dairy products. It should also be noted that these approaches are not effective when performing metagenomic shotgun sequencing, as there is no amplification step to eliminate the dye-bound DNA.

1.5.3.3.4 Outlook

Currently culture-independent, population-based, analysis is relatively expensive and, thus, further developments are needed to increase its relevance to the food industry. It is, however, becoming more accessible as a test method for companies to strategically analyze processing pipelines and end products, allowing development of targeted treatments and intervention strategies against persistent or troublesome microorganisms. To provide thorough and reproducible analysis of dairy powders in this fashion, it will be particularly important to arrive at a consensus regarding the standardized sample preparation, use of specific sequencing platforms and analysis methodologies to facilitate comparison across multiple investigations (Clooney *et al.*, 2016).

1.5.3.3.5 Further advances

More recently, culture independent, non-targeted DNA analysis has been employed to analyse dairy samples to gain insights into a number of factors influencing the dairy microbiota. 16S amplicon sequencing has been employed to determine the microbiota of bulk tank raw milk within and between farms (Skeie *et al.*, 2019), as well as the association between microbial populations with somatic cell counts and

bacterial counts (Rodrigues *et al.*, 2017). The impacts of weather (Li *et al.*, 2018), lactation stage, milk storage conditions (Doyle *et al.*, 2017a), as well as environment and farm management practices on the raw milk microbiota (Doyle *et al.*, 2017b; Doyle *et al.*, 2017a) and final cheese product (Fretin *et al.*, 2018) have been studied. The core and seasonal microbiota, the impact of transfer to a processing facility, and initial pooling and storage before product production have also been studied in this way (Kable *et al.*, 2016; Kable *et al.*, 2019). In addition, the temperature of feed to the ultrafiltration membrane has been shown to influence biofilm formation in processing facilities (Chamberland *et al.*, 2019). Shotgun metagenomic sequencing has also been employed to determine the microorganism responsible for a pink discolouration defect in cheese (Quigley *et al.*, 2016) as well as to characterise the airborne viromes in cheese production plants (Colombo *et al.*, 2018).

Taken together, these studies suggest that the microbial composition is influenced by a broad range of factors. These include the farm of origin, weather, season, lactation stage, environmental and farm management practices, storage conditions, transport, storage at the processing facility, milk feeding temperatures, and any environmental contamination from within the processing facility itself. The previous studies have used an Illumina-based sequencing approach, however a consensus view on the data is difficult due to confounding factors including different country of study, alternative milk production systems and seasons, varied DNA extraction kits and sequencing preparation, and different bioinformatics pipelines for data analysis.

Furthermore, there has been a limited focus on the role and detection of spore-forming bacteria throughout the milk processing chain and a lack of focus on dairy powders. A number of reviews have highlighted the benefits of sequencing techniques compared to current culture based technique used to identify sporeformers (Sadiq, Flint and He, 2018; Wells-Bennik, Driehuis and van Hijum, 2016). It has however been noted that the low concentrations of spores in dairy products may require pre-treatment, enrichment, or concentration, for microbiome analysis. This limits the rate at which sequencing can be implemented for routine screening in food industry (Wells-Bennik, Driehuis and van Hijum, 2016).

Overall, the broad study variability highlights a need for a standardised approach for routine microbiome analysis of processing facilities. This will enable milk processors to make informed decisions for production based on a greater understanding of the microbiota associated with the product.

1.6 Conclusions

Newer technologies have paved the way for an overhaul in the approaches taken to detect and enumerate of spore-forming bacteria in dairy powders. This can lead to a more accurate, high throughput system. Although the newer technologies themselves are not without their limitations, they are continuously improving. Optimization of these newer technologies could lead to their routine use, allowing development of improved targeted treatments and preventative measures in the powder processing industry.

1.7 References

Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V. and Polz, M. F. (2004) 'Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons', *Journal of Bacteriology,* 186(9), pp. 2629-2635. DOI: 10.1128/jb.186.9.2629-2635.2004

Andersson, A., Rönner, U. and Granum, P. E. (1995) 'What problems does the food industry have with the spore-forming pathogens *Bacillus cereus* and *Clostridium perfringens*?', *International Journal of Food Microbiology,* 28(2), pp. 145-155. DOI: 10.1016/0168-1605(95)00053-4

Ayers, S. H. and Mudge, C. S. (1920) 'Milk-powder agar for the determination of bacteria in milk', *J Bacteriol,* 5(6), pp. 565-88.

Barash, J. R., Hsia, J. K. and Arnon, S. S. (2010) 'Presence of soil-dwelling Clostridia in commercial powdered infant formulas', *J Pediatr,* 156(3), pp. 402-8. DOI: 10.1016/j.jpeds.2009.09.072

Beecher, D. J. and Wong, A. C. L. (1997) 'Tripartite hemolysin BL from *Bacillus cereus*: hemolytic analysis of component interactions and a model for its characteristic paradoxical zone phenomenon', *Journal of Biological Chemistry,* 272(1), pp. 233-239. DOI: 10.1074/jbc.272.1.233

Bengtsson-Palme, J., Boulund, F., Fick, J., Kristiansson, E. and Larsson, D. G. (2014) 'Shotgun metagenomics reveals a wide array of antibiotic resistance genes and

mobile elements in a polluted lake in India', *Front Microbiol,* 5, pp. 648. DOI: 10.3389/fmicb.2014.00648

Binga, E. K., Lasken, R. S. and Neufeld, J. D. (2008) 'Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology', *ISME J,* 2(3), pp. 233-41. DOI: 10.1038/ismej.2008.10

Black, M. W., VanderKelen, J., Montana, A., Dekhtyar, A., Neal, E., Goodman, A. and Kitts, C. L. (2014) 'Pyroprinting: a rapid and flexible genotypic fingerprinting method for typing bacterial strains', *J Microbiol Methods,* 105, pp. 121-9. DOI: 10.1016/j.mimet.2014.07.019

Boonstra, M., de Jong, I. G., Scholefield, G., Murray, H., Kuipers, O. P. and Veening, J. W. (2013) 'Spo0A regulates chromosome copy number during sporulation by directly binding to the origin of replication in *Bacillus subtilis*', *Mol Microbiol,* 87(4), pp. 925-38. DOI: 10.1111/mmi.12141

Bottone, E. J. (2010) '*Bacillus cereus*, a volatile human pathogen', *Clin Microbiol Rev,* 23(2), pp. 382-98. DOI: 10.1128/CMR.00073-09

Branda, S. S., Gonzalez-Pastor, J. E., Ben-Yehuda, S., Losick, R. and Kolter, R. (2001) 'Fruiting body formation by *Bacillus subtilis*', *Proc Natl Acad Sci U S A,* 98(20), pp. 11621-6. DOI: 10.1073/pnas.191384198

Bravo, A., Gómez, I., Porta, H., García-Gómez, B. I., Rodriguez-Almazan, C., Pardo, L. and Soberón, M. (2013) 'Evolution of *Bacillus thuringiensis* Cry toxins insecticidal

activity', *Microbial biotechnology,* 6(1), pp. 17-26. DOI: 10.1111/j.1751-7915.2012.00342.x

Brett, M. M., McLauchlin, J., Harris, A., O'Brien, S., Black, N., Forsyth, R. J., Roberts, D. and Bolton, F. J. (2005) 'A case of infant botulism with a possible link to infant formula milk powder: evidence for the presence of more than one strain of Clostridium botulinum in clinical specimens and food', *J Med Microbiol,* 54(Pt 8), pp. 769-76. DOI: 10.1099/jmm.0.46000-0

Brooks, J. P., Edwards, D. J., Harwich, M. D., Jr., Rivera, M. C., Fettweis, J. M., Serrano, M. G., Reris, R. A., Sheth, N. U., Huang, B., Girerd, P., Vaginal Microbiome, C., Strauss, J. F., 3rd, Jefferson, K. K. and Buck, G. A. (2015) 'The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies', *BMC Microbiol,* 15, pp. 66. DOI: 10.1186/s12866-015-0351-6

Brown, K. L. (2000) 'Control of bacterial spores', *British Medical Bulletin,* 56(1), pp. 158-171. DOI: 10.1258/0007142001902860

Buehner, K. P., Anand, S. and Djira, G. D. (2015) 'Prevalence of thermoduric bacteria and spores in nonfat dry milk powders of Midwest origin', *J Dairy Sci,* 98(5), pp. 2861-6. DOI: 10.3168/jds.2014-8822

Burgess, S. A., Lindsay, D. and Flint, S. H. (2010) 'Thermophilic bacilli and their importance in dairy processing', *Int J Food Microbiol,* 144(2), pp. 215-25. DOI: 10.1016/j.ijfoodmicro.2010.09.027

Case, R. J., Boucher, Y., Dahllof, I., Holmstrom, C., Doolittle, W. F. and Kjelleberg, S. (2007) 'Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies', *Appl Environ Microbiol,* 73(1), pp. 278-88. DOI: 10.1128/AEM.01177-06

Castiaux, V., Liu, X., Delbrassinne, L. and Mahillon, J. (2015) 'Is Cytotoxin K from *Bacillus cereus* a *bona fide* enterotoxin?', *Int J Food Microbiol,* 211, pp. 79-85. DOI: 10.1016/j.ijfoodmicro.2015.06.020

Cattani, F., Barth, V. C., Jr., Nasario, J. S., Ferreira, C. A. and Oliveira, S. D. (2016) 'Detection and quantification of viable *Bacillus cereus* group species in milk by propidium monoazide quantitative real-time PCR', *J Dairy Sci,* 99(4), pp. 2617-24. DOI: 10.3168/jds.2015-10019

Chamberland, J., Messier, T., Dugat-Bony, E., Lessard, M.-H., Labrie, S., Doyen, A. and Pouliot, Y. (2019) 'Influence of feed temperature to biofouling of ultrafiltration membrane during skim milk processing', *International Dairy Journal,* 93, pp. 99-105. DOI: 10.1016/j.idairyj.2019.02.005

Chang, Y. H., Shangkuan, Y. H., Lin, H. C. and Liu, H. W. (2003) 'PCR assay of the groEL gene for detection and differentiation of *Bacillus cereus* group cells', *Appl Environ Microbiol,* 69(8), pp. 4502-10. DOI: 10.1128/aem.69.8.4502-4510.2003

Checinska, A., Paszczynski, A. and Burbank, M. (2015) '*Bacillus* and other spore-forming genera: variations in responses and mechanisms for survival', *Annu Rev Food Sci Technol,* 6, pp. 351-69. DOI: 10.1146/annurev-food-030713-092332

Choma, C. and Granum, P. E. (2002) 'The enterotoxin T (BcET) from *Bacillus cereus* can probably not contribute to food poisoning', *FEMS Microbiology Letters,* 217(1), pp. 115-119. DOI: 10.1111/j.1574-6968.2002.tb11464.x

Clooney, A. G., Fouhy, F., Sleator, R. D., A, O. D., Stanton, C., Cotter, P. D. and Claesson, M. J. (2016) 'Comparing apples and oranges?: Next generation sequencing and its impact on microbiome analysis', *PLoS One,* 11(2), pp. e0148028. DOI: 10.1371/journal.pone.0148028

Colombo, S., Arioli, S., Gargari, G., Neri, E., Della Scala, G. and Mora, D. (2018) 'Characterization of airborne viromes in cheese production plants', *J Appl Microbiol*. DOI: 10.1111/jam.14046

Cui, Y., Liu, X., Dietrich, R., Martlbauer, E., Cao, J., Ding, S. and Zhu, K. (2016) 'Characterization of *Bacillus cereus* isolates from local dairy farms in China', *FEMS Microbiol Lett,* 363(12). DOI: 10.1093/femsle/fnw096

Dairy Australia (2015) *Australian dairy industry in focus 2015* (ISSN 2202-7467). Available at: https://www.dairyaustralia.com.au/-/media/dairyaustralia/documents/about-dairy-australia/key-publications/2015-in-focus.pdf?la=en&hash=C8CB36603249C6DC38C646B4F550EC9E92D9E99B. Last accessed 04/03/2020.

DCANZ (2016) *New Zealand milk production data report*. Available at: http://www.dcanz.com/. Last accessed 04/03/2020.

de Bourcy, C. F. A., De Vlaminck, I., Kanbar, J. N., Wang, J., Gawad, C. and Quake, S. R. (2014) 'A quantitative comparison of single-cell whole genome amplification methods', *PLoS ONE,* 9(8), pp. e105585. DOI: 10.1371/journal.pone.0105585

Deng, X., Li, Z. and Zhang, W. (2012) 'Transcriptome sequencing of *Salmonella enterica* serovar Enteritidis under desiccation and starvation stress in peanut oil', *Food Microbiol,* 30(1), pp. 311-5. DOI: 10.1016/j.fm.2011.11.001

Dhakal, R., Chauhan, K., Seale, R. B., Deeth, H. C., Pillidge, C. J., Powell, I. B., Craven, H. and Turner, M. S. (2013) 'Genotyping of dairy *Bacillus licheniformis* isolates by high resolution melt analysis of multiple variable number tandem repeat loci', *Food Microbiol,* 34(2), pp. 344-51. DOI: 10.1016/j.fm.2013.01.006

Doyle, C. J., Gleeson, D., Jordan, K., Beresford, T. P., Ross, R. P., Fitzgerald, G. F. and Cotter, P. D. (2015) 'Anaerobic sporeformers and their significance with respect to milk and dairy products', *Int J Food Microbiol,* 197, pp. 77-87. DOI: 10.1016/j.ijfoodmicro.2014.12.022

Doyle, C. J., Gleeson, D., O'Toole, P. W. and Cotter, P. D. (2017a) 'High-throughput metataxonomic characterization of the raw milk microbiota identifies changes reflecting lactation stage and storage conditions', *Int J Food Microbiol,* 255, pp. 1-6. DOI: 10.1016/j.ijfoodmicro.2017.05.019

Doyle, C. J., Gleeson, D., O'Toole, P. W. and Cotter, P. D. (2017b) 'Impacts of Seasonal Housing and Teat Preparation on Raw Milk Microbiota: a High-Throughput Sequencing Study', *Appl Environ Microbiol,* 83(2). DOI: 10.1128/AEM.02694-16

Driks, A. (2002) 'Maximum shields: the assembly and function of the bacterial spore coat', *Trends Microbiol,* 10(6), pp. 251-4. DOI: 10.1016/s0966-842x(02)02373-9

Durak, M. Z., Fromm, H. I., Huck, J. R., Zadoks, R. N. and Boor, K. J. (2006) 'Development of molecular typing methods for *Bacillus* spp. and *Paenibacillus* spp. isolated from fluid milk products', *Journal of Food Science,* 71(2), pp. M50-M56. DOI: 10.1111/j.1365-2621.2006.tb08907.x

Dutilh, B. E., Cassman, N., McNair, K., Sanchez, S. E., Silva, G. G., Boling, L., Barr, J. J., Speth, D. R., Seguritan, V., Aziz, R. K., Felts, B., Dinsdale, E. A., Mokili, J. L. and Edwards, R. A. (2014) 'A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes', *Nat Commun,* 5, pp. 4498. DOI: 10.1038/ncomms5498

Eisenberg, E. and Levanon, E. Y. (2013) 'Human housekeeping genes, revisited', *Trends Genet,* 29(10), pp. 569-74. DOI: 10.1016/j.tig.2013.05.010

European Commission (2005) 'Commission regulation (EC) No 2073/2005', *Official Journal of the European Union*.

Eurostat (2016) 'Cows'milk collection and products obtained - monthly data'. Available at: http://ec.europa.eu/eurostat/data/database. Last accessed 04/03/2020.

Faille, C., Benezech, T., Midelet-Bourdin, G., Lequette, Y., Clarisse, M., Ronse, G., Ronse, A. and Slomianny, C. (2014) 'Sporulation of *Bacillus* spp. within biofilms: a

potential source of contamination in food processing environments', *Food Microbiol,* 40, pp. 64-74. DOI: 10.1016/j.fm.2013.12.004

Feehery, G. R., Yigit, E., Oyola, S. O., Langhorst, B. W., Schmidt, V. T., Stewart, F. J., Dimalanta, E. T., Amaral-Zettler, L. A., Davis, T., Quail, M. A. and Pradhan, S. (2013) 'A method for selectively enriching microbial DNA from contaminating vertebrate host DNA', *PLoS ONE,* 8(10), pp. e76096. DOI: 10.1371/journal.pone.0076096

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., Owens, S., Gilbert, J. A., Wall, D. H. and Caporaso, J. G. (2012) 'Cross-biome metagenomic analyses of soil microbial communities and their functional attributes', *Proceedings of the National Academy of Sciences of the United States of America,* 109(52), pp. 21390-21395. DOI: 10.1073/pnas.1215210110

Filippidou, S., Junier, T., Wunderlin, T., Lo, C. C., Li, P. E., Chain, P. S. and Junier, P. (2015) 'Under-detection of endospore-forming Firmicutes in metagenomic data', *Comput Struct Biotechnol J,* 13, pp. 299-306. DOI: 10.1016/j.csbj.2015.04.002

Forghani, F., Langaee, T., Eskandari, M., Seo, K. H., Chung, M. J. and Oh, D. H. (2015) 'Rapid detection of viable *Bacillus cereus* emetic and enterotoxic strains in food by coupling propidium monoazide and multiplex PCR (PMA-mPCR)', *Food Control,* 55, pp. 151-157. DOI: 10.1016/j.foodcont.2015.02.049

Frank, J. F. and Yousef, A. E. (2004) 'Chapter 08 Tests for groups of microorganisms', in Frank, J.F. (ed.) *Standard Methods for the Examination of Dairy Products*: American Public Health Association. DOI: 10.2105/9780875530024ch08

Fretin, M., Martin, B., Rifa, E., Isabelle, V. M., Pomies, D., Ferlay, A., Montel, M. C. and Delbes, C. (2018) 'Bacterial community assembly from cow teat skin to ripened cheeses is influenced by grazing systems', *Sci Rep,* 8(1), pp. 200. DOI: 10.1038/s41598-017-18447-y

FSAI (2014)*: Guidelines for the interpretation of results of microbiological testing of ready-to-eat foods placed on the market (revision 1)*: Food Safety Authority of Ireland.

Gibbs, B. M. and Freame, B. (1965) 'Methods for the recovery of Clostridia from foods', *Journal of Applied Bacteriology,* 28(1), pp. 95-111. DOI: 10.1111/j.1365-2672.1965.tb02131.x

Gil, R., Silva, F. J., Pereto, J. and Moya, A. (2004) 'Determination of the core of a minimal bacterial gene set', *Microbiol Mol Biol Rev,* 68(3), pp. 518-37, table of contents. DOI: 10.1128/MMBR.68.3.518-537.2004

Gill, H. S., Rutherfurd, K. J. and Cross, M. L. (2001) 'Dietary probiotic supplementation enhances natural killer cell activity in the elderly: an investigation of age-related immunological changes', *Journal of Clinical Immunology,* 21(4), pp. 264-271. DOI: 10.1023/a:1010979225018

Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M. and Nelson, K. E. (2006) 'Metagenomic analysis of the human distal gut microbiome', *Science,* 312(5778), pp. 1355-9. DOI: 10.1126/science.1124234

Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B. and Chiodini, R. J. (2016) 'Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples', *Gut Pathog,* 8, pp. 24. DOI: 10.1186/s13099-016-0103-7

Gleeson, D., O'Connell, A. and Jordan, K. (2013) 'Review of potential sources and control of thermoduric bacteria in bulk-tank milk', *Irish Journal of Agricultural and Food Research,* 52(2), pp. 217-227.

Goodwin, S., McPherson, J. D. and McCombie, W. R. (2016) 'Coming of age: ten years of next-generation sequencing technologies', *Nat Rev Genet,* 17(6), pp. 333-51. DOI: 10.1038/nrg.2016.49

Gopal, N., Hill, C., Ross, P. R., Beresford, T. P., Fenelon, M. A. and Cotter, P. D. (2015) 'The prevalence and control of *Bacillus* and related spore-forming bacteria in the dairy industry', *Front Microbiol,* 6, pp. 1418. DOI: 10.3389/fmicb.2015.01418

Helgason, E., Okstad, O. A., Caugant, D. A., Johansen, H. A., Fouet, A., Mock, M., Hegna, I. and Kolsto, A. B. (2000) '*Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* - one species on the basis of genetic evidence', *Applied and Environmental Microbiology,* 66(6), pp. 2627-2630. DOI: 10.1128/Aem.66.6.2627-2630.2000

Heyndrickx, M. (2011) 'The importance of endospore-forming bacteria originating from soil for contamination of industrial food processing', *Applied and Environmental Soil Science,* 2011, pp. 1-11. DOI: 10.1155/2011/561975

Hoch, J. A. (1993) 'Regulation of the phosphorelay and the initiation of sporulation in *Bacillus subtilis*', *Annual Review of Microbiology,* 47(1), pp. 441-465. DOI: 10.1146/annurev.mi.47.100193.002301

Hong, H. A., Duc le, H. and Cutting, S. M. (2005) 'The use of bacterial spore formers as probiotics', *FEMS Microbiol Rev,* 29(4), pp. 813-35. DOI: 10.1016/j.femsre.2004.12.001

Horwood, P. F., Burgess, G. W. and Jane Oakey, H. (2004) 'Evidence for non-ribosomal peptide synthetase production of cereulide (the emetic toxin) in Bacillus cereus', *FEMS Microbiology Letters,* 236(2), pp. 319-324. DOI: 10.1111/j.1574-6968.2004.tb09664.x

Ivy, R. A., Ranieri, M. L., Martin, N. H., den Bakker, H. C., Xavier, B. M., Wiedmann, M. and Boor, K. J. (2012) 'Identification and characterization of psychrotolerant sporeformers associated with fluid milk production and processing', *Appl Environ Microbiol,* 78(6), pp. 1853-64. DOI: 10.1128/AEM.06536-11

Izadi, Z., Sheikh-Zeinoddin, M., Ensafi, A. A. and Soleimanian-Zad, S. (2016) 'Fabrication of an electrochemical DNA-based biosensor for *Bacillus cereus* detection in milk and infant formula', *Biosens Bioelectron,* 80, pp. 582-9. DOI: 10.1016/j.bios.2016.02.032

Janda, J. M. and Abbott, S. L. (2002) 'Bacterial identification for publication: when is enough enough?', *Journal of Clinical Microbiology,* 40(6), pp. 1887-1891. DOI: 10.1128/jcm.40.6.1887-1891.2002

Johnson, E. A., Tepp, W. H., Bradshaw, M., Gilbert, R. J., Cook, P. E. and McIntosh, E. D. G. (2005) 'Characterization of Clostridium botulinum Strains Associated with an Infant Botulism Case in the United Kingdom', *Journal of Clinical Microbiology,* 43(6), pp. 2602-2607. DOI: 10.1128/jcm.43.6.2602-2607.2005

Jung, J. Y., Lee, S. H., Kim, J. M., Park, M. S., Bae, J. W., Hahn, Y., Madsen, E. L. and Jeon, C. O. (2011) 'Metagenomic analysis of kimchi, a traditional Korean fermented food', *Appl Environ Microbiol,* 77(7), pp. 2264-74. DOI: 10.1128/AEM.02157-10

Kable, M. E., Srisengfa, Y., Laird, M., Zaragoza, J., McLeod, J., Heidenreich, J. and Marco, M. L. (2016) 'The Core and Seasonal Microbiota of Raw Bovine Milk in Tanker Trucks and the Impact of Transfer to a Milk Processing Facility', *MBio,* 7(4). DOI: 10.1128/mBio.00836-16

Kable, M. E., Srisengfa, Y., Xue, Z., Coates, L. C. and Marco, M. L. (2019) 'Viable and total bacterial populations undergo equipment- and time-dependent shifts during milk processing', *Appl Environ Microbiol*. DOI: 10.1128/AEM.00270-19

Karam, M. C., Gaiani, C., Hosri, C., Burgain, J. and Scher, J. (2013) 'Effect of dairy powders fortification on yogurt textural and sensorial properties: a review', *J Dairy Res,* 80(4), pp. 400-9. DOI: 10.1017/S0022029913000514

Kent, D. J., Chauhan, K., Boor, K. J., Wiedmann, M. and Martin, N. H. (2016) 'Spore test parameters matter: Mesophilic and thermophilic spore counts detected in raw milk and dairy powders differ significantly by test method', *J Dairy Sci*. DOI: 10.3168/jds.2015-10283

Kent, R. M., Fitzgerald, G. F., Hill, C., Stanton, C. and Ross, R. P. (2015) 'Novel approaches to improve the intrinsic microbiological safety of powdered infant milk formula', *Nutrients,* 7(2), pp. 1217-44. DOI: 10.3390/nu7021217

Lagrange, V., Whitsett, D. and Burris, C. (2015) 'Global market for dairy proteins', *J Food Sci,* 80 Suppl 1, pp. A16-22. DOI: 10.1111/1750-3841.12801

Lai, E. M., Phadke, N. D., Kachman, M. T., Giorno, R., Vazquez, S., Vazquez, J. A., Maddock, J. R. and Driks, A. (2003) 'Proteomic analysis of the spore coats of *Bacillus subtilis* and *Bacillus anthracis*', *Journal of Bacteriology,* 185(4), pp. 1443-1454. DOI: 10.1128/jb.185.4.1443-1454.2003

Lee, J. L. and Levin, R. E. (2009) 'A comparative study of the ability of EMA and PMA to distinguish viable from heat killed mixed bacterial flora from fish fillets', *J Microbiol Methods,* 76(1), pp. 93-6. DOI: 10.1016/j.mimet.2008.08.008

Leonard, S. R., Mammel, M. K., Lacher, D. W. and Elkins, C. A. (2015) 'Application of metagenomic sequencing to food safety: detection of Shiga toxin-producing *Escherichia coli* on fresh bagged spinach', *Appl Environ Microbiol,* 81(23), pp. 8183-91. DOI: 10.1128/AEM.02601-15

Li, A. G., Burggraf, L. W. and Xing, Y. (2016) 'Nanomechanical characterization of *Bacillus anthracis* spores by atomic force microscopy', *Appl Environ Microbiol,* 82(10), pp. 2988-99. DOI: 10.1128/AEM.00431-16

Li, N., Wang, Y., You, C., Ren, J., Chen, W., Zheng, H. and Liu, Z. (2018) 'Variation in Raw Milk Microbiota Throughout 12 Months and the Impact of Weather Conditions', *Sci Rep,* 8(1), pp. 2371. DOI: 10.1038/s41598-018-20862-8

Lindback, T., Fagerlund, A., Rodland, M. S. and Granum, P. E. (2004) 'Characterization of the *Bacillus cereus* Nhe enterotoxin', *Microbiology,* 150(Pt 12), pp. 3959-67. DOI: 10.1099/mic.0.27359-0

Liu, Y., Lai, Q., Dong, C., Sun, F., Wang, L., Li, G. and Shao, Z. (2013) 'Phylogenetic diversity of the *Bacillus pumilus* group and the marine ecotype revealed by multilocus sequence analysis', *PLoS ONE,* 8(11), pp. e80097. DOI: 10.1371/journal.pone.0080097

Liu, Y., Lai, Q., Goker, M., Meier-Kolthoff, J. P., Wang, M., Sun, Y., Wang, L. and Shao, Z. (2015) 'Genomic insights into the taxonomic status of the *Bacillus cereus* group', *Sci Rep,* 5, pp. 14082. DOI: 10.1038/srep14082

Lund, T., De Buyser, M. L. and Granum, P. E. (2000) 'A new cytotoxin from *Bacillus cereus* that may cause necrotic enteritis', *Mol Microbiol,* 38(2), pp. 254-61.

Lund, T. and Granum, P. E. (1996) 'Characterisation of a non-haemolytic enterotoxin complex from *Bacillus cereus* isolated after a foodborne outbreak', *FEMS Microbiology Letters,* 141(2-3), pp. 151-156. DOI: 10.1111/j.1574-6968.1996.tb08377.x

Luu, S., Cruz-Mora, J., Setlow, B., Feeherry, F. E., Doona, C. J. and Setlow, P. (2015) 'The effects of heat activation on *Bacillus* spore germination, with nutrients or

under high pressure, with or without various germination proteins', *Appl Environ Microbiol,* 81(8), pp. 2927-38. DOI: 10.1128/AEM.00193-15

Madoui, M. A., Engelen, S., Cruaud, C., Belser, C., Bertrand, L., Alberti, A., Lemainque, A., Wincker, P. and Aury, J. M. (2015) 'Genome assembly using nanopore-guided long and error-free DNA reads', *BMC Genomics,* 16, pp. 327. DOI: 10.1186/s12864-015-1519-z

Matz, L. L., Beaman, T. C. and Gerhardt, P. (1970) 'Chemical composition of exosporium from spores of *Bacillus cereus*', *Journal of Bacteriology,* 101(1), pp. 196-201.

Meeske, A. J., Rodrigues, C. D., Brady, J., Lim, H. C., Bernhardt, T. G. and Rudner, D. Z. (2016) 'High-throughput genetic screens identify a large and diverse collection of new sporulation genes in *Bacillus subtilis*', *PLoS Biol,* 14(1), pp. e1002341. DOI: 10.1371/journal.pbio.1002341

Mertens, K., Freund, L., Schmoock, G., Hansel, C., Melzer, F. and Elschner, M. C. (2014) 'Comparative evaluation of eleven commercial DNA extraction kits for real-time PCR detection of *Bacillus anthracis* spores in spiked dairy samples', *Int J Food Microbiol,* 170, pp. 29-37. DOI: 10.1016/j.ijfoodmicro.2013.10.022

Miller, R. A., Kent, D. J., Watterson, M. J., Boor, K. J., Martin, N. H. and Wiedmann, M. (2015) 'Spore populations among bulk tank raw milk and dairy powders are significantly different', *J Dairy Sci,* 98(12), pp. 8492-504. DOI: 10.3168/jds.2015-9943

Minami, J., Soejima, T., Yaeshima, T. and Iwatsuki, K. (2012) 'Direct real-time PCR with ethidium monoazide: a method for the rapid detection of viable *Cronobacter sakazakii* in powdered infant formula', *J Food Prot,* 75(9), pp. 1572-9. DOI: 10.4315/0362-028X.JFP-12-015

Molle, V., Fujita, M., Jensen, S. T., Eichenberger, P., González-Pastor, J. E., Liu, J. S. and Losick, R. (2003) 'The Spo0A regulon of *Bacillus subtilis*', *Molecular Microbiology,* 50(5), pp. 1683-1701. DOI: 10.1046/j.1365-2958.2003.03818.x

Nicholson, W. L., Munakata, N., Horneck, G., Melosh, H. J. and Setlow, P. (2000) 'Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments', *Microbiology and Molecular Biology Reviews,* 64(3), pp. 548-572. DOI: 10.1128/mmbr.64.3.548-572.2000

Nocker, A. and Camper, A. K. (2006) 'Selective removal of DNA from dead cells of mixed bacterial communities by use of ethidium monoazide', *Appl Environ Microbiol,* 72(3), pp. 1997-2004. DOI: 10.1128/AEM.72.3.1997-2004.2006

Nocker, A., Cheung, C. Y. and Camper, A. K. (2006) 'Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells', *J Microbiol Methods,* 67(2), pp. 310-20. DOI: 10.1016/j.mimet.2006.04.015

Okstad, O. A. and Kolsto, A. B. (2011) 'Genomics of *Bacillus* species', *Genomics of Foodborne Bacterial Pathogens*, pp. 29-53. DOI: 10.1007/978-1-4419-7686-4_2

Pantoja, J. C., Reinemann, D. J. and Ruegg, P. L. (2011) 'Factors associated with coliform count in unpasteurized bulk milk', *J Dairy Sci,* 94(6), pp. 2680-91. DOI: 10.3168/jds.2010-3721

Parente, E., Cocolin, L., De Filippis, F., Zotta, T., Ferrocino, I., O'Sullivan, O., Neviani, E., De Angelis, M., Cotter, P. D. and Ercolini, D. (2016) 'FoodMicrobionet: A database for the visualisation and exploration of food bacterial communities based on network analysis', *Int J Food Microbiol,* 219, pp. 28-37. DOI: 10.1016/j.ijfoodmicro.2015.12.001

Pennacchia, C., Breeuwer, P. and Meyer, R. (2014) 'Development of a Multiplex-PCR assay for the rapid identification of Geobacillus stearothermophilus and Anoxybacillus flavithermus', *Food Microbiol,* 43, pp. 41-9. DOI: 10.1016/j.fm.2014.05.002

Pinheiro, L. B., Coleman, V. A., Hindson, C. M., Herrmann, J., Hindson, B. J., Bhat, S. and Emslie, K. R. (2012) 'Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification', *Analytical Chemistry,* 84(2), pp. 1003-1011. DOI: 10.1021/ac202578x

Pompeo, F., Foulquier, E. and Galinier, A. (2016) 'Impact of serine/threonine protein kinases on the regulation of sporulation in *Bacillus subtilis*', *Front Microbiol,* 7, pp. 568. DOI: 10.3389/fmicb.2016.00568

Popham, D. L. (2002) 'Specialized peptidoglycan of the bacterial endospore: the inner wall of the lockbox', *Cell Mol Life Sci,* 59(3), pp. 426-33. DOI: 10.1007/s00018-002-8435-5

Porcellato, D., Narvhus, J. and Skeie, S. B. (2016) 'Detection and quantification of *Bacillus cereus* group in milk by droplet digital PCR', *J Microbiol Methods*. DOI: 10.1016/j.mimet.2016.05.012

Poretsky, R., Rodriguez, R. L., Luo, C., Tsementzi, D. and Konstantinidis, K. T. (2014) 'Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics', *PLoS One,* 9(4), pp. e93827. DOI: 10.1371/journal.pone.0093827

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P. and Gu, Y. (2012) 'A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers', *BMC Genomics,* 13(1), pp. 1-13. DOI: 10.1186/1471-2164-13-341

Quigley, L., O'Sullivan, D. J., Daly, D., O'Sullivan, O., Burdikova, Z., Vana, R., Beresford, T. P., Ross, R. P., Fitzgerald, G. F., McSweeney, P. L., Giblin, L., Sheehan, J. J. and Cotter, P. D. (2016) 'Thermus and the Pink Discoloration Defect in Cheese', *mSystems,* 1(3). DOI: 10.1128/mSystems.00023-16

Radnedge, L., Agron, P. G., Hill, K. K., Jackson, P. J., Ticknor, L. O., Keim, P. and Andersen, G. L. (2003) 'Genome differences that distinguish *Bacillus anthracis* from *Bacillus cereus* and *Bacillus thuringiensis*', *Applied and Environmental Microbiology,* 69(5), pp. 2755-2764. DOI: 10.1128/Aem.69.5.2755-2764.2003

Rasko, D. A., Altherr, M. R., Han, C. S. and Ravel, J. (2005) 'Genomics of the *Bacillus cereus* group of organisms', *FEMS Microbiol Rev,* 29(2), pp. 303-29. DOI: 10.1016/j.femsre.2004.12.005

Reginensi, S. M., Gonzalez, M. J., Olivera, J. A., Sosa, M., Juliano, P. and Bermudez, J. (2011) 'RAPD-based screening for spore-forming bacterial populations in Uruguayan commercial powdered milk', *Int J Food Microbiol,* 148(1), pp. 36-41. DOI: 10.1016/j.ijfoodmicro.2011.04.020

Reyes, J. E., Bastias, J. M., Gutierrez, M. R. and Rodriguez Mde, L. (2007) 'Prevalence of *Bacillus cereus* in dried milk products used by Chilean school feeding program', *Food Microbiol,* 24(1), pp. 1-6. DOI: 10.1016/j.fm.2006.04.004

Rodrigues, M. X., Lima, S. F., Canniatti-Brazaca, S. G. and Bicalho, R. C. (2017) 'The microbiome of bulk tank milk: Characterization and associations with somatic cell count and bacterial count', *J Dairy Sci*. DOI: 10.3168/jds.2016-11540

Ronimus, R. S., Parker, L. E. and Morgan, H. W. (1997) 'The utilization of RAPD-PCR for identifying thermophilic and mesophilic *Bacillus* species', *FEMS Microbiol Lett,* 147(1), pp. 75-9. DOI: 10.1111/j.1574-6968.1997.tb10223.x

Ronimus, R. S., Parker, L. E., Turner, N., Poudel, S., Ruckert, A. and Morgan, H. W. (2003) 'A RAPD-based comparison of thermophilic Bacilli from milk powders', *Int J Food Microbiol,* 85(1-2), pp. 45-61. DOI: 10.1016/s0168-1605(02)00480-4

Ruckert, A., Ronimus, R. S. and Morgan, H. W. (2004) 'A RAPD-based survey of thermophilic Bacilli in milk powders from different countries', *Int J Food Microbiol,* 96(3), pp. 263-72. DOI: 10.1016/j.ijfoodmicro.2004.03.020

Rudi, K., Moen, B., Dromtorp, S. M. and Holck, A. L. (2005) 'Use of ethidium monoazide and PCR in combination for quantification of viable and dead cells in

complex samples', *Appl Environ Microbiol,* 71(2), pp. 1018-24. DOI: 10.1128/AEM.71.2.1018-1024.2005

Rueckert, A., Ronimus, R. S. and Morgan, H. W. (2005) 'Development of a rapid detection and enumeration method for thermophilic Bacilli in milk powders', *J Microbiol Methods,* 60(2), pp. 155-67. DOI: 10.1016/j.mimet.2004.09.008

Sadiq, F. A., Flint, S. and He, G. (2018) 'Microbiota of milk powders and the heat resistance and spoilage potential of aerobic spore-forming bacteria', *International Dairy Journal,* 85, pp. 159-168. DOI: 10.1016/j.idairyj.2018.06.003

Sadiq, F. A., Li, Y., Liu, T., Flint, S., Zhang, G. and He, G. (2016) 'A RAPD based study revealing a previously unreported wide range of mesophilic and thermophilic spore formers associated with milk powders in China', *Int J Food Microbiol,* 217, pp. 200-8. DOI: 10.1016/j.ijfoodmicro.2015.10.030

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J. and Walker, A. W. (2014) 'Reagent and laboratory contamination can critically impact sequence-based microbiome analyses', *BMC Biology,* 12(1), pp. 1-12. DOI: 10.1186/s12915-014-0087-z

Schellenberg, J., Links, M. G., Hill, J. E., Hemmingsen, S. M., Peters, G. A. and Dumonceaux, T. J. (2011) 'Pyrosequencing of chaperonin-60 (cpn60) amplicons as a means of determining microbial community composition', in Kwon, M.Y. and Ricke, C.S. (eds.) *High-Throughput Next Generation Sequencing: Methods and Applications*. Totowa, NJ: Humana Press, pp. 143-158. DOI: 10.1007/978-1-61779-089-8_10

Schellenberg, J. J., Paramel Jayaprakash, T., Withana Gamage, N., Patterson, M. H., Vaneechoutte, M. and Hill, J. E. (2016) '*Gardnerella vaginalis* subgroups defined by *cpn*60 sequencing and sialidase activity in isolates from Canada, Belgium and Kenya', *PLoS ONE,* 11(1), pp. e0146510. DOI: 10.1371/journal.pone.0146510

Schirmer, A., Gadkari, R., Reeves, C. D., Ibrahim, F., DeLong, E. F. and Hutchinson, C. R. (2005) 'Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*', *Appl Environ Microbiol,* 71(8), pp. 4840-9. DOI: 10.1128/AEM.71.8.4840-4849.2005

Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D. R. and Dean, D. H. (1998) '*Bacillus thuringiensis* and its pesticidal crystal proteins', *Microbiology and Molecular Biology Reviews,* 62(3), pp. 775-806.

Seale, R. B., Dhakal, R., Chauhan, K., Craven, H. M., Deeth, H. C., Pillidge, C. J., Powell, I. B. and Turner, M. S. (2012) 'Genotyping of present-day and historical *Geobacillus* species isolates from milk powders by high-resolution melt analysis of multiple variable-number tandem-repeat loci', *Appl Environ Microbiol,* 78(19), pp. 7090-7. DOI: 10.1128/AEM.01817-12

Seinige, D., Krischek, C., Klein, G. and Kehrenberg, C. (2014) 'Comparative analysis and limitations of ethidium monoazide and propidium monoazide treatments for the differentiation of viable and nonviable *Campylobacter* cells', *Appl Environ Microbiol,* 80(7), pp. 2186-92. DOI: 10.1128/AEM.03962-13

Sella, S. R., Vandenberghe, L. P. and Soccol, C. R. (2014) 'Life cycle and spore resistance of spore-forming *Bacillus atrophaeus*', *Microbiol Res,* 169(12), pp. 931-9. DOI: 10.1016/j.micres.2014.05.001

Setlow, P. (2003) 'Spore germination', *Current Opinion in Microbiology,* 6(6), pp. 550-556. DOI: 10.1016/j.mib.2003.10.001

Setlow, P. (2006) 'Spores of *Bacillus subtilis*: their resistance to and killing by radiation, heat and chemicals', *J Appl Microbiol,* 101(3), pp. 514-25. DOI: 10.1111/j.1365-2672.2005.02736.x

Setlow, P. (2007) 'I will survive: DNA protection in bacterial spores', *Trends Microbiol,* 15(4), pp. 172-80. DOI: 10.1016/j.tim.2007.02.004

Setlow, P. (2014) 'Germination of spores of *Bacillus* species: what we know and do not know', *J Bacteriol,* 196(7), pp. 1297-305. DOI: 10.1128/JB.01455-13

Sharma, A., Jana, A. H. and Chavan, R. S. (2012) 'Functionality of milk powders and milk-based powders for end use applications-a review', *Comprehensive Reviews in Food Science and Food Safety,* 11(5), pp. 518-528. DOI: 10.1111/j.1541-4337.2012.00199.x

Sharpton, T. J. (2014) 'An introduction to the analysis of shotgun metagenomic data', *Front Plant Sci,* 5, pp. 209. DOI: 10.3389/fpls.2014.00209

Skeie, S. B., Haland, M., Thorsen, I. M., Narvhus, J. and Porcellato, D. (2019) 'Bulk tank raw milk microbiota differs within and between farms: A moving goalpost challenging quality control', *J Dairy Sci*. DOI: 10.3168/jds.2017-14083

Sobel, J. (2005) 'Botulism', *Clinical Infectious Diseases,* 41(8), pp. 1167-1173. DOI: 10.1086/444507

Soejima, T., Minami, J.-i., Xiao, J.-z. and Abe, F. (2016) 'Innovative use of platinum compounds to selectively detect live microorganisms by polymerase chain reaction', *Biotechnology and Bioengineering,* 113(2), pp. 301-310. DOI: 10.1002/bit.25711

Standards (2004)*: ISO 7932:2004*.

Steffen, M. M., Li, Z., Effler, T. C., Hauser, L. J., Boyer, G. L. and Wilhelm, S. W. (2012) 'Comparative metagenomics of toxic freshwater Cyanobacteria bloom communities on two continents', *PLoS ONE,* 7(8), pp. e44002. DOI: 10.1371/journal.pone.0044002

Stewart, G. C. (2015) 'The exosporium layer of bacterial spores: a connection to the environment and the infected host', *Microbiol Mol Biol Rev,* 79(4), pp. 437-57. DOI: 10.1128/MMBR.00050-15

Sunde, E. P., Setlow, P., Hederstedt, L. and Halle, B. (2009) 'The physical state of water in bacterial spores', *Proc Natl Acad Sci U S A,* 106(46), pp. 19334-9. DOI: 10.1073/pnas.0908712106

Sylvestre, P., Couture-Tosi, E. and Mock, M. (2002) 'A collagen-like surface glycoprotein is a structural component of the *Bacillus anthracis* exosporium', *Mol Microbiol,* 45(1), pp. 169-78. DOI: 10.1046/j.1365-2958.2000.03000.x

Tallent, S. M., Kotewicz, K. M., Strain, E. A. and Bennett, R. W. (2012) 'Efficient isolation and identification of *Bacillus cereus* group', *J AOAC Int,* 95(2), pp. 446-51. DOI: 10.5740/jaoacint.11-251

Tamir, H. and Gilvarg, C. (1966) 'Density gradient centrifugation for the separation of sporulating forms of bacteria', *Journal of Biological Chemistry,* 241(5), pp. 1085-1090.

Toh, M., Moffitt, M. C., Henrichsen, L., Raftery, M., Barrow, K., Cox, J. M., Marquis, C. P. and Neilan, B. A. (2004) 'Cereulide, the emetic toxin of *Bacillus cereus*, is putatively a product of nonribosomal peptide synthesis', *J Appl Microbiol,* 97(5), pp. 992-1000. DOI: 10.1111/j.1365-2672.2004.02381.x

Tran, S. L., Guillemet, E., Gohar, M., Lereclus, D. and Ramarao, N. (2010) 'CwpFM (EntFM) is a *Bacillus cereus* potential cell wall peptidase implicated in adhesion, biofilm formation, and virulence', *J Bacteriol,* 192(10), pp. 2638-42. DOI: 10.1128/JB.01315-09

Trmcic, A., Martin, N. H., Boor, K. J. and Wiedmann, M. (2015) 'A standard bacterial isolate set for research on contemporary dairy spoilage', *J Dairy Sci,* 98(8), pp. 5806-17. DOI: 10.3168/jds.2015-9490

USDA (1968)*: United States standards for grades of edible dry casein (acid)*: United States Department of Agriculture Agricultural Marketing Service. Available at: https://www.ams.usda.gov/sites/default/files/media/Edible%20Dry%20Casein%20%28Acid%29%20Standards.pdf. Last accessed 04/03/2020.

USDA (1984)*: United States standards for grades of nonfat dry milk (roller process)*: United States Department of Agriculture Agricultural Marketing Service. Available at: https://www.ams.usda.gov/sites/default/files/media/Nonfat_Dry_Milk_%28Roller_Process%29_Standard%5B1%5D.pdf. Last accessed 04/03/2020.

USDA (2000)*: United states standards for dry whey*: United States Department of Agriculture Agricultural Marketing Service. Available at: https://www.ams.usda.gov/sites/default/files/media/Dry_Whey_Standard%5B1%5D.pdf. Last accessed 04/03/2020.

USDA (2001a)*: United states standards for grades of dry buttermilk and dry buttermilk product*: United States Department of Agriculture Agricultural Marketing Service. Available at: https://www.ams.usda.gov/sites/default/files/media/Dry_Buttermilk_and_Dry_Buttermilk_Product_Standard%5B1%5D.pdf. Last accessed 04/03/2020.

USDA (2001b)*: United states standards for grades of dry whole milk*: United States Department of Agriculture Agricultural Marketing Service. Available at: https://www.ams.usda.gov/sites/default/files/media/Dry_Whole_Milk_Standard%5B1%5D.pdf. Last accessed 04/03/2020.

USDA (2001c)*: United states standards for grades of nonfat dry milk (spray process)*: United States Department of Agriculture Agricultural Marketing Service. Available at:

https://www.ams.usda.gov/sites/default/files/media/Nonfat_Dry_Milk_%28Spray_Process%29_Standard%5B1%5D.pdf. Last accessed 04/03/2020.

USDA (2013)*: United states standards for instant nonfat dry milk*: United States Department of Agriculture Agricultural Marketing Service. Available at: https://www.ams.usda.gov/sites/default/files/media/Instant_Nonfat_Dry_Milk_Standard%5B1%5D.pdf. Last accessed 04/03/2020.

USDA (2016) *Milk production* (ISSN: 1949-1557). Available at:https://downloads.usda.library.cornell.edu/usda-esmis/files/h989r321c/rf55z895w/th83m0669/MilkProd-09-20-2016.pdf. Last accessed 04/03/2020.

VanderKelen, J. J., Mitchell, R. D., Laubscher, A., Black, M. W., Goodman, A. L., Montana, A. K., Dekhtyar, A. M., Jimenez-Flores, R. and Kitts, C. L. (2016) 'Short communication: Typing and tracking Bacillaceae in raw milk and milk powder using pyroprinting', *J Dairy Sci,* 99(1), pp. 146-51. DOI: 10.3168/jds.2015-9656

Vangay, P., Fugett, E. B., Sun, Q. and Wiedmann, M. (2013) 'Food microbe tracker: a web-based tool for storage and comparison of food-associated microbes', *J Food Prot,* 76(2), pp. 283-94. DOI: 10.4315/0362-028X.JFP-12-276

Vincent, A. T. and Charette, S. J. (2015) 'Who qualifies to be a bioinformatician?', *Frontiers in Genetics,* 6. DOI: 10.3389/fgene.2015.00164

Watterson, M. J., Kent, D. J., Boor, K. J., Wiedmann, M. and Martin, N. H. (2014) 'Evaluation of dairy powder products implicates thermophilic sporeformers as the

primary organisms of interest', *J Dairy Sci,* 97(4), pp. 2487-97. DOI: 10.3168/jds.2013-7363

Weenk, G. H., van den Brink, J. A., Struijk, C. B. and Mossel, D. A. A. (1995) 'Modified methods for the enumeration of spores of mesophilic *Clostridium* species in dried foods', *International Journal of Food Microbiology,* 27(2), pp. 185-200. DOI: 10.1016/0168-1605(94)00164-2

Wehrle, E., Moravek, M., Dietrich, R., Burk, C., Didier, A. and Martlbauer, E. (2009) 'Comparison of multiplex PCR, enzyme immunoassay and cell culture methods for the detection of enterotoxinogenic *Bacillus cereus*', *J Microbiol Methods,* 78(3), pp. 265-70. DOI: 10.1016/j.mimet.2009.06.013

Weinmaier, T., Probst, A. J., La Duc, M. T., Ciobanu, D., Cheng, J. F., Ivanova, N., Rattei, T. and Vaishampayan, P. (2015) 'A viability-linked metagenomic analysis of cleanroom environments: eukarya, prokaryotes, and viruses', *Microbiome,* 3, pp. 62. DOI: 10.1186/s40168-015-0129-y

Wells-Bennik, M. H., Eijlander, R. T., den Besten, H. M., Berendsen, E. M., Warda, A. K., Krawczyk, A. O., Nierop Groot, M. N., Xiao, Y., Zwietering, M. H., Kuipers, O. P. and Abee, T. (2016) 'Bacterial spores in food: survival, emergence, and outgrowth', *Annu Rev Food Sci Technol,* 7, pp. 457-82. DOI: 10.1146/annurev-food-041715-033144

Wells-Bennik, M. H. J., Driehuis, F. and van Hijum, S. A. F. T. (2016) 'Prospects for improved control of dairy-relevant sporeformers using -omics technologies', *Current Opinion in Food Science,* 10, pp. 38-44. DOI: 10.1016/j.cofs.2016.08.005

Wielinga, P. R., de Heer, L., de Groot, A., Hamidjaja, R. A., Bruggeman, G., Jordan, K. and van Rotterdam, B. J. (2011) 'Evaluation of DNA extraction methods for *Bacillus anthracis* spores spiked to food and feed matrices at biosafety level 3 conditions', *Int J Food Microbiol,* 150(2-3), pp. 122-7. DOI: 10.1016/j.ijfoodmicro.2011.07.023

Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. and Tingey, S. V. (1990) 'DNA polymorphisms amplified by arbitrary primers are useful as genetic-markers', *Nucleic Acids Research,* 18(22), pp. 6531-6535. DOI: 10.1093/nar/18.22.6531

Wilson, W. J. and Blair, E. M. M. V. (1924) 'The application of a sulphite-glucose-irobn agar medium to the quantitative estimation of *B. welchii* and other reducing bacteria in water supplies', *The Journal of Pathology and Bacteriology,* 27(1), pp. 119-121. DOI: 10.1002/path.1700270114

Wongputtisin, P., Khanongnuch, C., Kongbuntad, W., Niamsup, P., Lumyong, S. and Sarkar, P. K. (2014) 'Use of *Bacillus subtilis* isolates from Tua-nao towards nutritional improvement of soya bean hull for monogastric feed application', *Letters in Applied Microbiology,* 59(3), pp. 328-333. DOI: 10.1111/lam.12279

Wu, G. P., Chen, S. H. and Levin, R. E. (2015) 'Application of ethidium bromide monoazide for quantification of viable and dead cells of *Salmonella enterica* by real-time loop-mediated isothermal amplification', *J Microbiol Methods,* 117, pp. 41-8. DOI: 10.1016/j.mimet.2015.07.012

Wunderlin, T., Junier, T., Paul, C., Jeanneret, N. and Junier, P. (2016) 'Physical isolation of endospores from environmental samples by targeted lysis of vegetative cells', *J Vis Exp,* (107), pp. e53411. DOI: 10.3791/53411

Yokouchi, H., Fukuoka, Y., Mukoyama, D., Calugay, R., Takeyama, H. and Matsunaga, T. (2006) 'Whole-metagenome amplification of a microbial community associated with scleractinian coral by multiple displacement amplification using phi29 polymerase', *Environ Microbiol,* 8(7), pp. 1155-63. DOI: 10.1111/j.1462-2920.2006.01005.x

Zhang, Z., Feng, L., Xu, H., Liu, C., Shah, N. P. and Wei, H. (2016) 'Detection of viable enterotoxin-producing *Bacillus cereus* and analysis of toxigenicity from ready-to-eat foods and infant formula milk powder by multiplex PCR', *J Dairy Sci,* 99(2), pp. 1047-55. DOI: 10.3168/jds.2015-10147

Zhang, Z., Liu, W., Xu, H., Aguilar, Z. P., Shah, N. P. and Wei, H. (2015) 'Propidium monoazide combined with real-time PCR for selective detection of viable *Staphylococcus aureus* in milk powder and meat products', *J Dairy Sci,* 98(3), pp. 1625-33. DOI: 10.3168/jds.2014-8938

Zheng, L., Abhyankar, W., Ouwerling, N., Dekker, H. L., van Veen, H., van der Wel, N. N., Roseboom, W., de Koning, L. J., Brul, S. and de Koster, C. G. (2016) '*Bacillus subtilis* spore inner membrane proteome', *J Proteome Res,* 15(2), pp. 585-94. DOI: 10.1021/acs.jproteome.5b00976

Zhu, K., Holzel, C. S., Cui, Y., Mayer, R., Wang, Y., Dietrich, R., Didier, A., Bassitta, R., Martlbauer, E. and Ding, S. (2016a) 'Probiotic *Bacillus cereus* strains, a potential risk for public health in China', *Front Microbiol,* 7, pp. 718. DOI: 10.3389/fmicb.2016.00718

Zhu, L., He, J., Cao, X., Huang, K., Luo, Y. and Xu, W. (2016b) 'Development of a double-antibody sandwich ELISA for rapid detection of *Bacillus Cereus* in food', *Sci Rep,* 6, pp. 16092. DOI: 10.1038/srep16092

Chapter 2. Compositional-based assessment of the impact of ultra-high temperature processing on the composition of thermophilic sporeformers in whey powder

Authors: Aoife J. McHugh, Fang Li, Karen Hunt, John T. Tobin, Conor Feehily, Colin Hill and Paul D. Cotter

Contributions: Candidate performed sample collection, enrichment, DNA extraction, sequencing library preparation, bioinformatic and statistical analysis.

Candidate drafted and edited chapter.

FL provided the powder and organised UHT trials.

FL performed culture-based analysis.

KH and JTT supervised FL.

CF, CH and PDC supervised candidate, revised and edited chapter.

2.1 Abstract

Thermophilic sporeformers selected for by heat and drying treatments during dairy powder production have the ability to stay dormant in the desiccated powder for extended periods. These sporeformers are of concern for dairy processors as they have the potential to cause spoilage or illness. Traditional culture-based techniques are widely applied to determine counts of thermophilic spore-forming bacteria, however there is no accurate culture-based means of simultaneously identifying all sporeformers present. Without an accurate determination of sporeformer composition, it is not possible to fully appreciate the associated risks or to assess the success with which targeted treatments, such as ultra-high temperature (UHT) processing, can remove undesirable taxa. Here we applied 16S rDNA amplicon sequencing of metagenomic DNA as part of a proof-of-concept study. The main aim was to assess the relative impact of different UHT treatments in a pilot-scale plant on the dominant spore-forming populations present in a batch of whey powder. Ethidium bromide monoazide (EMA)-treatment was employed to distinguish between DNA sourced from living and dead bacteria. *Geobacillus*, *Bacillus,* and *Brevibacillus* were the most dominant thermophilic genera identified, and their relative abundances varied depending on the treatment temperature employed. This study represents a promising approach to harness the potential of high throughput DNA sequencing to assess the microbial composition of products from a dairy processing facility. It provides a platform to facilitate the further application of this technology to investigate the factors that influence the microbiology of food processing facilities.

2.2 Introduction

Spore-forming bacteria are a concern to dairy processors due to their ability to withstand the high temperatures and resultant low moisture content associated with dairy powder production as well as their ability to form biofilms (Faille *et al.*, 2014), cause spoilage (Burgess, Lindsay and Flint, 2010; Sadiq, Flint and He, 2018) or contribute to disease (Gopal *et al.*, 2015; Majed *et al.*, 2016). Some sporeformers are also capable of forming heat stable enzymes (Sadiq, Flint and He, 2018) and toxins (Lucking *et al.*, 2013; Majed *et al.*, 2016) and can be resistant to cleaning in place techniques used in production lines (Zou and Liu, 2018). Furthermore, even though the majority of sporeformers do not cause spoilage or illness and are ubiquitous in the environment, their presence in food products can be viewed as an indicator of poor hygiene (Burgess, Lindsay and Flint, 2010).

Thermophilic sporeformers are of great concern, and are more prevalent than mesophilic and psychrotrophic sporeformers in non-fat dry milk powder and whey powders (Watterson *et al.*, 2014) and in powder processing lines (Cho *et al.*, 2018). Consequently, there is a need for interventions in powder processing facilities to prevent contamination by these microorganisms (Ortuzar *et al.*, 2018; Miller *et al.*, 2015). Within dairy powders, *Bacillus* and *Geobacillus* sp. have been identified as the dominant spore-forming genera (Miller *et al.*, 2015; McHugh *et al.*, 2018). There are a variety of different culture-based approaches that can be employed to quantify levels of spore-forming bacteria in powders, but with no clear widely applied international guidelines (Miller *et al.*, 2015). Nonetheless, stringent, customer-enforced, count specifications for sporeformers in dairy powders are

often applied (Watterson *et al.*, 2014). Ultra-high temperature (UHT) processing can be applied to reduce thermophilic sporeformers. A wide variety of temperatures and thermal cycles can be employed in different processing runs. Here we carry out an investigation within a small scale production plant that operates within the same constraints of and is indicative of an actual industrial processing run, to determine the effect of UHT processing on thermophilic sporeformer composition. The microbial content was investigated using 16S rDNA amplicon sequencing on a proof-of-concept basis to assess the usefulness of this approach for investigations of the products of dairy processing.

2.3 Materials and methods

2.3.1 Sample pasteurisation and enrichment

A single batch of whey powder was used for this trial. The whey powder was handled aseptically and reconstituted at a concentration of 10% w/v in sterile water for each trial. Initial culturing was performed to determine spore load. Mesophilic sporeformers and high heat resistant (HHR) sporeformers were enumerated following isolation on plate count skimmed milk agar (PCSMA; Merck). Briefly, reconstituted whey powder solution was heated to 80°C for 10 min prior to serial dilution and pour plating with incubation at 30°C for 48 h to select for mesophilic sporeformers. HHR thermophilic sporeformers were selected for by heat treating solutions to 100°C for 30 min, cooling immediately on ice prior to serial dilution and pour plating with incubation at 55°C for 48 h. Bacterial counts were expressed as CFU ml^{-1} of reconstituted whey protein powder. Spreading colonies were counted as single colonies if less than one quarter of the agar surface was covered; if more than one quarter of the agar surface was covered, the result was discarded. UHT trials were performed in triplicate to produce biological replicates (Trial 1 (T1), T2, T3). Reconstituted whey was subject to UHT treatments of 70-120°C for 30 sec in increments of 10°C with a resting temperature of 60°C. Specifically, whey was heated to and kept at 60°C before subjecting it to 70°C, 80°C, 90°C, 100°C, 110°C or 120°C for 30 sec and collecting 100 ml aseptically after each treatment for enrichment and DNA extraction with a further aliquot taken for culturing. Heat treatment was applied using an indirect tubular UHT processing pilot plant (MicroThermics, NC, USA), consisting of five components including a pre-heating

tubular heat exchanger, final heating tubular heat exchanger, holding tubes and two cooling units. Samples were taken and incubated at 55°C for 24 h to select for thermophilic bacteria, as initial culturing results indicated higher levels of thermophilic HHR sporeformers than mesophilic sporeformers present. Samples were then cultured as previously described for HHR spore-forming bacteria. Two 50 ml samples were taken for each temperature treatment. Separately, as a control, one 50 ml sample of reconstituted whey powder (5 g) was incubated at 55°C for 24 h only in order to reveal the thermophilic populations present in the absence of UHT treatment (T0).

2.3.2 DNA extraction and library preparation

Samples were centrifuged at 900 x g (Zhang *et al.*, 2015) for 20 min in order to pellet any food debris. The pellet was discarded and supernatant centrifuged at 4500 x g for 20 min in order to pellet bacterial cells. Pellets were then washed in 1 ml sterile ¼ Ringers and centrifuged at 13,000 x g for 2 min. Washing was repeated twice more, centrifuging for 1 min at 13,000 x g. Each sample was resuspended in ¼ ringers to 2 ml volume. Ethidium bromide monoazide (EMA) (Sigma-Aldrich, E2028-5MG), suspended in ethanol, at a concentration of 5 mg ml^{-1}, was added to one of the two samples from each pasteurisation temperature at a concentration of 10 µg ml^{-1} (Seinige *et al.*, 2014). The same volume of ethanol was added to the corresponding sample pair as a control. Samples were incubated in the dark for 5 min, before exposure to 70 Watt HQI light for 10 min at 20 cm while keeping the samples on ice. Samples were centrifuged at 12000 x g for 5 min, supernatant was discarded and pellet resuspended in 150 µl of fresh lysozyme from chicken egg

white at 45 mg ml^{-1} (Sigma-Aldrich, L4919-5G). Samples were incubated at 37°C for 30 min followed by centrifugation at 13000 x g for 1 min and supernatant removed. Finally, DNA was extracted using the PowerFood® Microbial DNA Isolation Kit (Cambio Ltd., 21000-100), including an alternative lysis step for difficult to lyse cells (McHugh *et al.*, 2018). DNA was eluted in 60 µl of 10 mM Tris-HCl and stored at -20°C. DNA was also extracted from T0 sample in the same way as samples without EMA treatment.

DNA was quantified and checked for quality using Qubit® dsDNA HS Assay Kit Qubit dsDNA High Sensitivity Assay kit (BioSciences Ltd., Dublin Q32854) as well as running on 2 µl on 1% (w/v) agarose gel. All samples with enough DNA were normalized by diluting to 5 ng μl^{-1}. 16S V3-V4 amplicon libraries were prepared by following the Illumina 16S Metagenomic sequencing library preparation guidelines with a few modifications. Briefly, the amplicon PCR was performed in triplicate with 35 cycles. Amplicons were visualised by running 2 µl on 1% (w/v) agarose gel before pooling each triplicate prior to purification with 0.8 X Ampure XP (Labplan Ltd., A63880). Samples were indexed and purified according to Illumina protocol and stored at -20°C. Indexed DNA was quantified using Qubit dsDNA High Sensitivity Assay kit, before diluting to 20 nM, pooling, and cleaning up 1:1 DNA pool:Ampure XP using the previously used protocol and eluting in the same volume. The 16S rDNA V3-V4 amplicon library was sequenced on an Illumina MiSeq using 2 x 250 V2 kit at the Teagasc Food Research Centre Sequencing facility, Fermoy.

2.3.3 Bioinformatic analysis

Forward and reverse reads were joined using FLASH (fast length adjustment of short reads to improve genome assemblies) (Magoc and Salzberg, 2011). Paired end reads were further processed by quality filtering based on quality score of 25 and removing mismatched barcodes and sequences below length thresholds by QIIME (Caporaso *et al.*, 2010b). A total of 4,858,040 reads were generated after filtering, with an average of 147,213.3 per sample and median 142,666 reads per sample. USEARCH v7 (64-bit) (Edgar, 2010) was utilised for removing noisy data, chimera detections, and clustering into operational taxonomic units (OTUs) at 97% identity. OTUs were aligned using PyNAST (python nearest alignment space termination) (Caporaso *et al.*, 2010a) and taxonomy was assigned using BLAST (Altschul *et al.*, 1990) against the SILVA SSURef database release 119 (Quast *et al.*, 2013). QIIME data was further analysed using Phyloseq in R (McMurdie and Holmes, 2013), and plots generated using Phyloseq, ggplot2 (Wickham, 2009), and cowplot. Alpha diversity was analysed in R using Phyloseq with Observed, Chao1, ACE, Shannon, Simpson and Fisher diversity metrics analysed. Bray-Curtis beta diversity was analysed in R using Phyloseq. Pairwise Wilcoxon matched pairs signed rank test was used with Benjamini-Hochberg corrections for multiple comparisons to determine if EMA or UHT treatment had an effect on sample composition and alpha diversity subject to UHT trials (ignored T0 in all analysis, and 120°C sample for temperature statistics as not enough replicates). Adonis from the vegan R package was used to determine significant differences in beta diversity due to EMA and temperature treatment (without T0 and 120°C samples).

2.4 Results

2.4.1 Proportions of spore-forming genera differ in response to different UHT temperatures

Initial culturing of the whey powder that was the focus of this study highlighted the presence of 1.7 log CFU ml^{-1} mesophilic sporeformers and 2.65 log CFU ml^{-1} thermophilic HHR sporeformers. After enriching reconstituted whey powder for thermophiles at 55°C, culture-based analysis showed decreasing average HHR thermophilic sporeformer counts from 2.56 log CFU ml^{-1} to 1.74 log CFU ml^{-1} with increasing UHT treatments between 70°C-90°C. However, average count then began to increase to 2.06 log CFU ml^{-1} with increasing temperatures from 90°C-110°C, although standard deviation also increased across these samples. The 120°C treatment yielded an average HHR thermophilic sporeformer count of 0.125 log CFU ml^{-1} (Supplemental Figure 2.1).

Following DNA extraction, the 120°C treated samples from T2 and T3 did not contain sufficient DNA to sequence. High throughput 16S rRNA sequencing revealed that *Geobacillus*, *Bacillus*, *Brevibacillus* and *Aneurinibacillus* were the dominant genera present in the remaining samples. The genus *Geobacillus* was present at the highest relative abundance in the T0 sample (Figure 2.1A). However, following UHT processing over a range or temperatures between 70°C and 100°C, the relative abundance of *Geobacillus* decreased, with *Bacillus* becoming the most dominant genus. When 110°C UHT treatment was employed, *Geobacillus*, *Bacillus* and *Brevibacillus* were present in equivalent proportions. In contrast, following the 120°C UHT treatment, *Geobacillus* was dominant (Figure 2.1A). A combination of

average genus level taxonomic classification with average CFU count data revealed levels of *Geobacillus* in 120°C sample were negligible by culturing (Supplemental Figure 2.2). When EMA treatment was employed to remove DNA signal from dead microorganisms, significant differences in the taxonomic composition outputs were noted (70-120°C treated samples). Proportions of *Bacillus* were significantly higher in samples treated with EMA compared to those not treated with EMA (Figure 2.1B, $p < 0.001$). Relative abundances of *Brevibacillus*, *Aneuribacillus*, *Streptococcus*, and *Burkholderia* were significantly lower (Figure 2.1B, $p < 0.01$) in samples treated with EMA compared to those not treated. No significant difference in relative abundance of the genus *Geobacillus* was noted due to EMA use (Figure 2.1B).

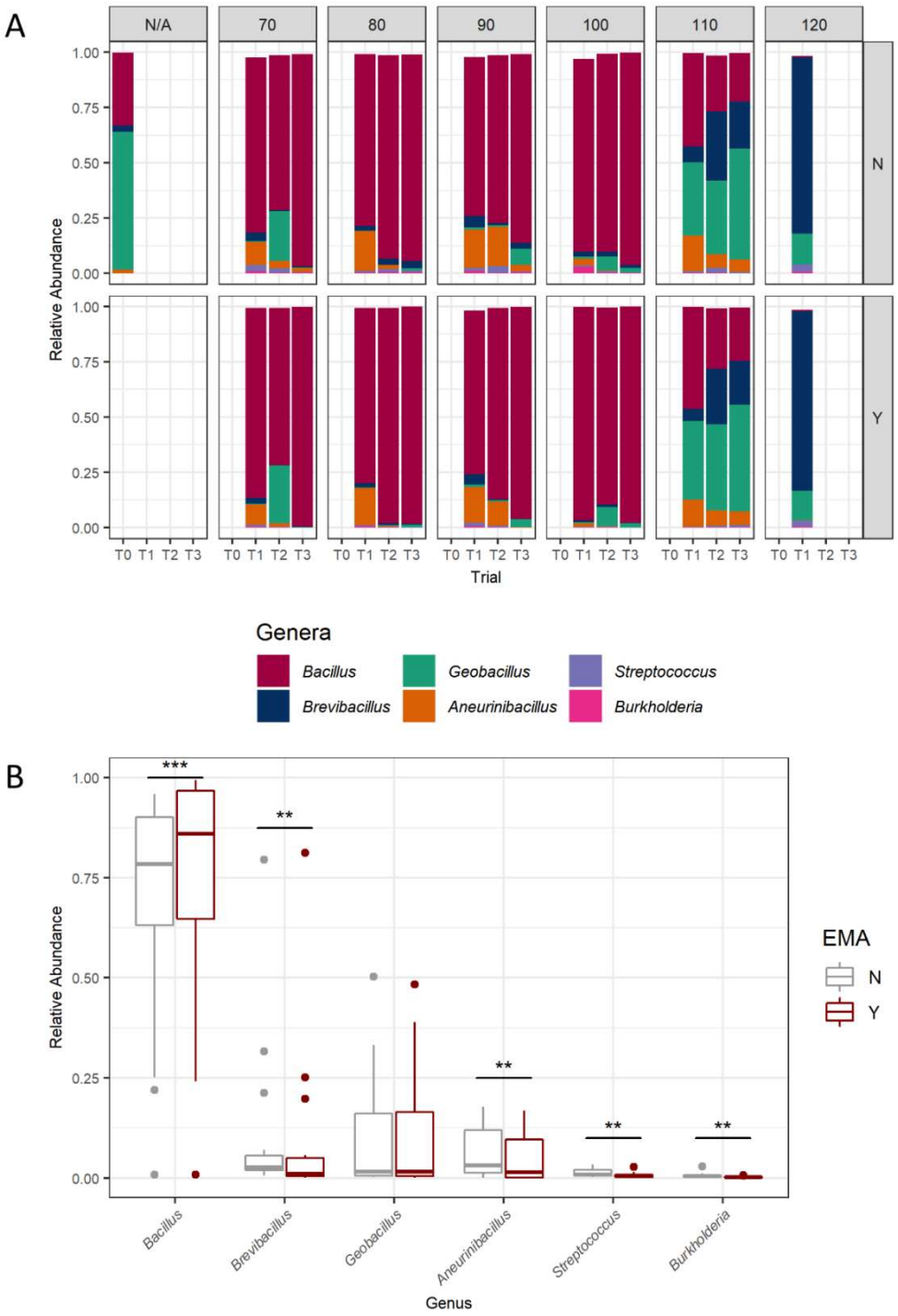


Figure 2.1. Relative abundance of genera per enriched sample.

A. Relative abundance of each genera per enriched sample. Plot is split by UHT treatment temperature (70-120) vertically, EMA use (Y/N) horizontally and trial along the x-axis. N/A represents T0 sample that was not subject to any UHT treatment.

[Figure 2.1 continued]

B. Relative abundance of genera grouped by EMA use. Significant differences in the relative abundance of genera between samples extracted using EMA (Y) and not (N) are shown (*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

2.4.2 EMA influences alpha diversity metrics when combined with enrichment

Significant differences in the alpha diversity metrics Shannon and Simpson (Figure 2.2A) were noted due to EMA use, with lower median alpha diversity in samples which had EMA applied (Figure 2.2A, $p < 0.001$). No significant differences were noted in Observed, Chao1, ACE, or Fisher alpha diversity metrics (Figure 2.2A). No significant difference in alpha diversity was noted due to temperature treatment (70 - 110°C) (Figure 2.2B).

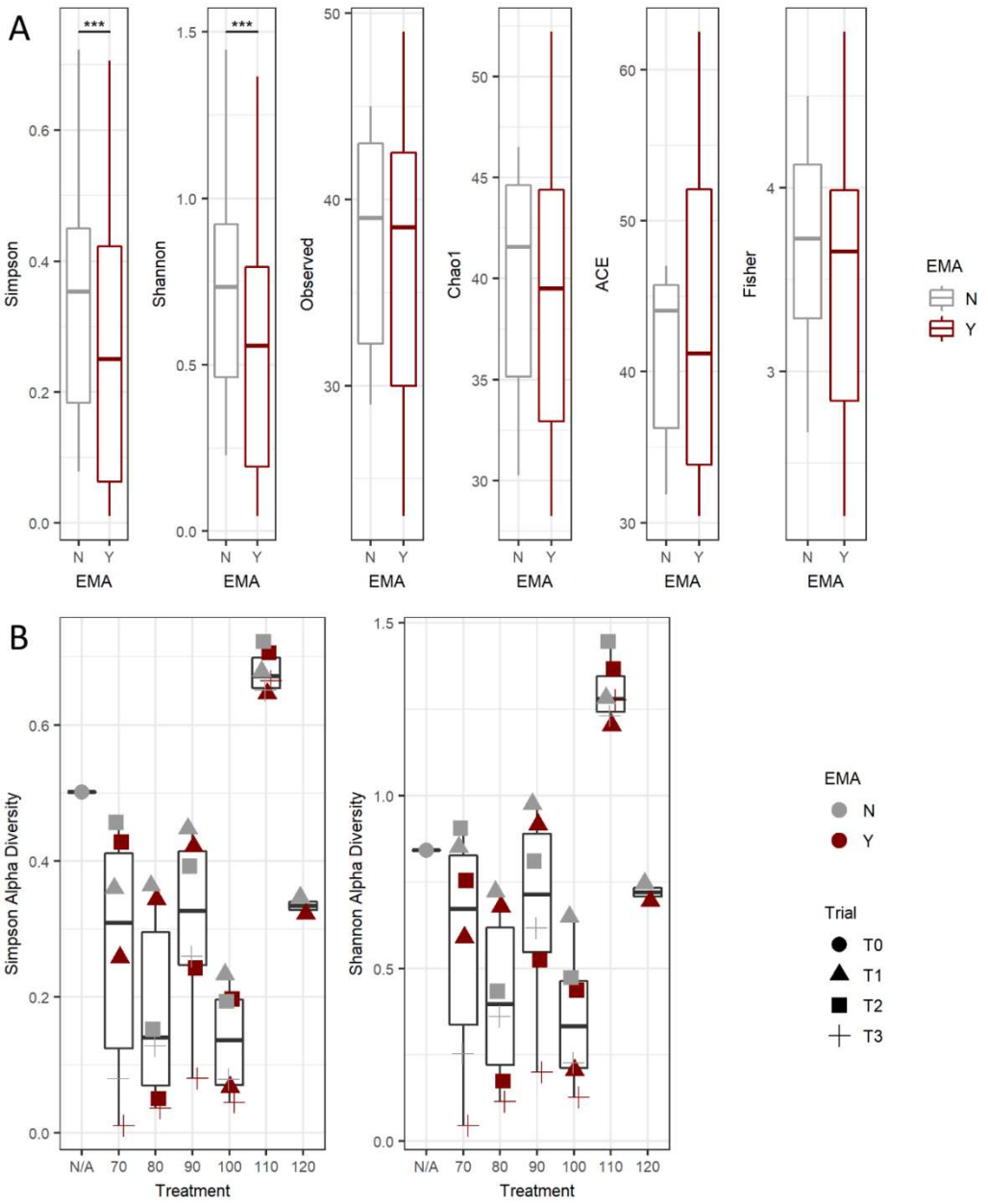


Figure 2.2. Alpha diversity of samples.

A. Alpha diversity of samples grouped by EMA use. Simpson, Shannon, Observed, Chao1, ACE, and Fisher alpha diversity metrics are shown. Significant differences in the alpha diversity between samples extracted using EMA (Y) and not (N) are shown (*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

[Figure 2.2 continued]

B. Simpson and Shannon alpha diversity of each sample. Simpson and Shannon alpha diversity of each sample is shown, split by UHT treatment temperature along the x-axis, points are coloured by EMA treatment (N/Y), and shaped by trial. N/A represents T0 sample that was not subject to any UHT treatment.

2.4.3 Temperature treatment causes significant differences in the beta diversity of samples between samples.

Significant differences were noted in the beta-diversity between samples, with temperature treatment explaining significant variation ($p < 0.001$) in the dissimilarity between samples (vegan:adonis, seed set at 999). These differences are of particular note as all the samples originated from one bag of powder. Following the removal of T0 and 120°C samples from statistical analysis (as there were not enough replicates), 55.12% variance was still explained by temperature treatment ($p < 0.001$). Additionally, trial (T1, T2, T3) accounted for 17.22% of the variation between samples ($p < 0.001$) (Figure 2.3). No significant difference between samples based on EMA use was noted. The diversity between samples treated at the same time with the same UHT processing was lowest in pairs treated with and without EMA (Figure 2.3).

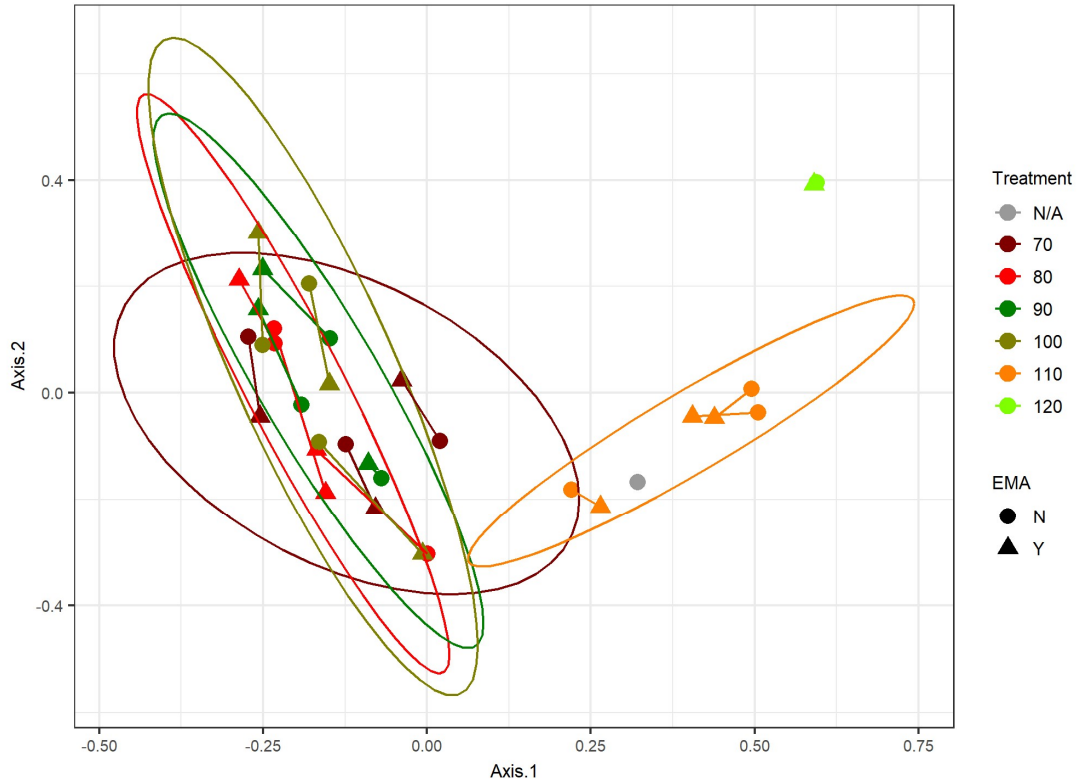


Figure 2.3. Bray Curtis MDS beta diversity analysis of samples.

Samples are coloured by UHT treatment with ellipses surrounding these groups when there are enough replicates (70-110). Shape of the points signifies EMA use. Bars joining two points signify paired samples extracted at the same time (with/without EMA) at each trial (T1, T2, T3). N/A represents T0 sample that was not subject to any UHT treatment.

2.5 Discussion

HHR thermophilic sporeformers were the dominant sporeformers present in the whey powder in this study as revealed by culture-based analysis. Unsurprisingly spore-forming bacteria were detected in all samples by 16S rRNA amplicon sequencing. The relative abundance of spore-forming genera differed depending on whether or not UHT processing was performed prior to enrichment and the UHT temperature used. *Bacillus*, *Brevibacillus* and *Geobacillus* were the dominant genera. This was similar to previous studies that identified *Bacillus* and *Geobacillus* as the dominant genera in dairy powders (Miller *et al.*, 2015; McHugh *et al.*, 2018). *Bacillus* is a common spore-forming genus found in dairy and environmental samples. *Geobacillus* is a thermophilic sporeformer and produces highly heat resistant spores. Like *Bacillus* and *Geobacillus*, *Brevibacillus* too has been previously identified in dairy (Gopal *et al.*, 2015). No significant differences were noted in taxonomic classification due to UHT treatment, though some non-significant variations were seen. Culture-based analysis revealed decreasing counts of thermophilic HHR sporeformer CFUs with increasing UHT treatments between 70-90°C. UHT treatments from 90-100°C resulted in increasing counts of HHR thermophilic CFU, though with considerable standard deviations. Overall this suggests the germination rates were affected due to different UHT treatments. After 120°C UHT treatment an average of 0.125 log CFU ml^{-1} was noted, which reflects the low DNA yield obtained from these samples. These analyses provide some insights into limits of detection for both culture, and culture independent methods.

Notably, EMA-treatment did impact on the taxonomic profile generated, with higher proportions of the already dominant *Bacillus* genus in samples treated with EMA and lower proportions of sub-dominant genera. This is presumably as a consequence of the removal of DNA from dead cells corresponding to such sub-dominant genera. EMA was used rather than propidium monoazide (PMA). There were a number of reasons for this. These powders had undergone heat treatment and drying in their production, as well as further heat treatments in the UHT processing in these trials. This lead to the presumption that many viable cells that were present in the raw milk ingredients were killed, as well as the bovine host cells being compromised, resulting in the ratio of dead cells far outweighing the surviving viable spore-forming bacteria targeted in this trial. Previously it was shown that lower concentrations of EMA (10 μg ml^{-1}) are more capable of supressing high ratios of dead DNA signal, than higher concentrations of the more expensive PMA (Seinige *et al.*, 2014). Low concentrations also suppress the dyes ability to penetrate live cells (Seinige *et al.*, 2014), which was previously reported to be an issue with EMA use (Nocker, Cheung and Camper, 2006).

Lower diversity within samples treated with EMA are due to decreases in the apparent relative abundance of sub-dominant genera, and/or increases in the relative abundances of the dominant *Bacillus* genus, reducing evenness. Shannon and Simpson diversity metrics incorporate evenness and these are the metrics whose values are significantly affected by EMA use. Observed species, Chao1, ACE and Fisher diversity results were not significantly impacted by EMA use, these metrics are less reliant on evenness and more reliant on compositional differences.

As EMA reduced the apparent relative abundance of sub-dominant genera and didn't eliminate their signal, the alpha diversity values from these compositionally influenced metrics were not significantly impacted.

Beta diversity analysis showed no significant difference between samples based on EMA use. However, there were significant differences between samples based on UHT treatment. This was shown by the fluctuating relative abundances of genera based on temperature treatment. Significantly greater variances between temperature treatments was noteworthy as all samples originate from one bag of powder. Pairs of samples from each treatment temperature and trial (each sample with and without EMA) were least dissimilar, showed no significant differences due to EMA use and had the lowest variances between them.

2.6 Conclusion

Temperature treatment had an effect on the microbial diversity of spore-forming bacteria in the whey powder tested. This was reflected in fluctuating relative abundances of genera depending on temperature treatment. Overall EMA use had an effect on the alpha diversity values within samples due to significantly increased relative abundance in *Bacillus*, the genus with highest relative abundance, and decreased abundance of low relative abundance genera with its use. EMA use resulted in decreased evenness and significantly decreased alpha diversity in metrics that incorporated evenness.

The UHT treatment employed in this study in a small-scale production plant that operates within the same constraints of, and is indicative of, an actual industrial processing run. Results highlight that processing temperatures differentially impact populations of sporeformers in dairy powder processing. This study represents a subset of possible UHT treatments and their impact on a small subset of populations. Ultimately, any number of treatments could be applied in incorporating these ingredients to products and these may all differentially impact the composition of spore-forming bacteria present. An understanding of the counts of these bacteria but also their composition will allow negation of risks and allow safe and economical storage as well as incorporation of these powders into a wide range of products.

Overall, different UHT treatment of powder post production differentially impact the composition of thermophilic sporeformers present. High throughput DNA sequencing is a viable tool to assess the microbial composition of products from a

dairy processing facility. Implementation of these methods could provide a greater understanding and allow investigation of the factors that influence the microbiology of food processing facilities.

2.7 References

Altschul, S., Gish, W., Miller, W., Myers, E. and Lipman, D. (1990) 'Basic local alignment search tool', *J Mol Biol,* 215. DOI: 10.1016/s0022-2836(05)80360-2

Burgess, S. A., Lindsay, D. and Flint, S. H. (2010) 'Thermophilic bacilli and their importance in dairy processing', *Int J Food Microbiol,* 144(2), pp. 215-25. DOI: 10.1016/j.ijfoodmicro.2010.09.027

Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L. and Knight, R. (2010a) 'PyNAST: a flexible tool for aligning sequences to a template alignment', *Bioinformatics,* 26(2), pp. 266-267. DOI: 10.1093/bioinformatics/btp636

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. and Knight, R. (2010b) 'QIIME allows analysis of high-throughput community sequencing data', *Nature Methods,* 7, pp. 335. DOI: 10.1038/nmeth.f.303

Cho, T. J., Kim, H. W., Kim, N. H., Park, S. M., Kwon, J. I., Kim, Y. J., Lee, K. W. and Rhee, M. S. (2018) 'New insights into the thermophilic spore-formers in powdered infant formula: Implications of changes in microbial composition during manufacture', *Food Control,* 92, pp. 464-470. DOI: 10.1016/j.foodcont.2018.05.036

Edgar, R. C. (2010) 'Search and clustering orders of magnitude faster than BLAST', *Bioinformatics,* 26(19), pp. 2460-2461. DOI: 10.1093/bioinformatics/btq461

Faille, C., Benezech, T., Midelet-Bourdin, G., Lequette, Y., Clarisse, M., Ronse, G., Ronse, A. and Slomianny, C. (2014) 'Sporulation of *Bacillus* spp. within biofilms: a potential source of contamination in food processing environments', *Food Microbiol,* 40, pp. 64-74. DOI: 10.1016/j.fm.2013.12.004

Gopal, N., Hill, C., Ross, P. R., Beresford, T. P., Fenelon, M. A. and Cotter, P. D. (2015) 'The prevalence and control of *Bacillus* and related spore-forming bacteria in the dairy industry', *Front Microbiol,* 6, pp. 1418. DOI: 10.3389/fmicb.2015.01418

Lucking, G., Stoeckel, M., Atamer, Z., Hinrichs, J. and Ehling-Schulz, M. (2013) 'Characterization of aerobic spore-forming bacteria associated with industrial dairy processing environments and product spoilage', *Int J Food Microbiol,* 166(2), pp. 270-9. DOI: 10.1016/j.ijfoodmicro.2013.07.004

Magoc, T. and Salzberg, S. L. (2011) 'FLASH: fast length adjustment of short reads to improve genome assemblies', *Bioinformatics,* 27(21), pp. 2957-63. DOI: 10.1093/bioinformatics/btr507

Majed, R., Faille, C., Kallassy, M. and Gohar, M. (2016) 'Bacillus cereus Biofilms-Same, Only Different', *Front Microbiol,* 7, pp. 1054. DOI: 10.3389/fmicb.2016.01054

McHugh, A. J., Feehily, C., Tobin, J. T., Fenelon, M. A., Hill, C. and Cotter, P. D. (2018) 'Mesophilic Sporeformers Identified in Whey Powder by Using Shotgun

Metagenomic Sequencing', *Appl Environ Microbiol,* 84(20). DOI: 10.1128/AEM.01305-18

McMurdie, P. J. and Holmes, S. (2013) 'phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data', *PLOS ONE,* 8(4), pp. e61217. DOI: 10.1371/journal.pone.0061217

Miller, R. A., Kent, D. J., Watterson, M. J., Boor, K. J., Martin, N. H. and Wiedmann, M. (2015) 'Spore populations among bulk tank raw milk and dairy powders are significantly different', *J Dairy Sci,* 98(12), pp. 8492-504. DOI: 10.3168/jds.2015-9943

Nocker, A., Cheung, C. Y. and Camper, A. K. (2006) 'Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells', *J Microbiol Methods,* 67(2), pp. 310-20. DOI: 10.1128/AEM.72.3.1997-2004.2006

Ortuzar, J., Martinez, B., Bianchini, A., Stratton, J., Rupnow, J. and Wang, B. (2018) 'Quantifying changes in spore-forming bacteria contamination along the milk production chain from farm to packaged pasteurized milk using systematic review and meta-analysis', *Food Control,* 86, pp. 319-331. DOI: 10.1016/j.foodcont.2017.11.038

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F. O. (2013) 'The SILVA ribosomal RNA gene database project: improved data processing and web-based tools', *Nucleic Acids Research,* 41(D1), pp. D590-D596. DOI: 10.1093/nar/gks1219

Sadiq, F. A., Flint, S. and He, G. (2018) 'Microbiota of milk powders and the heat resistance and spoilage potential of aerobic spore-forming bacteria', *International Dairy Journal,* 85, pp. 159-168. DOI: 10.1016/j.idairyj.2018.06.003

Seinige, D., Krischek, C., Klein, G. and Kehrenberg, C. (2014) 'Comparative analysis and limitations of ethidium monoazide and propidium monoazide treatments for the differentiation of viable and nonviable *Campylobacter* cells', *Appl Environ Microbiol,* 80(7), pp. 2186-92. DOI: 10.1128/AEM.03962-13

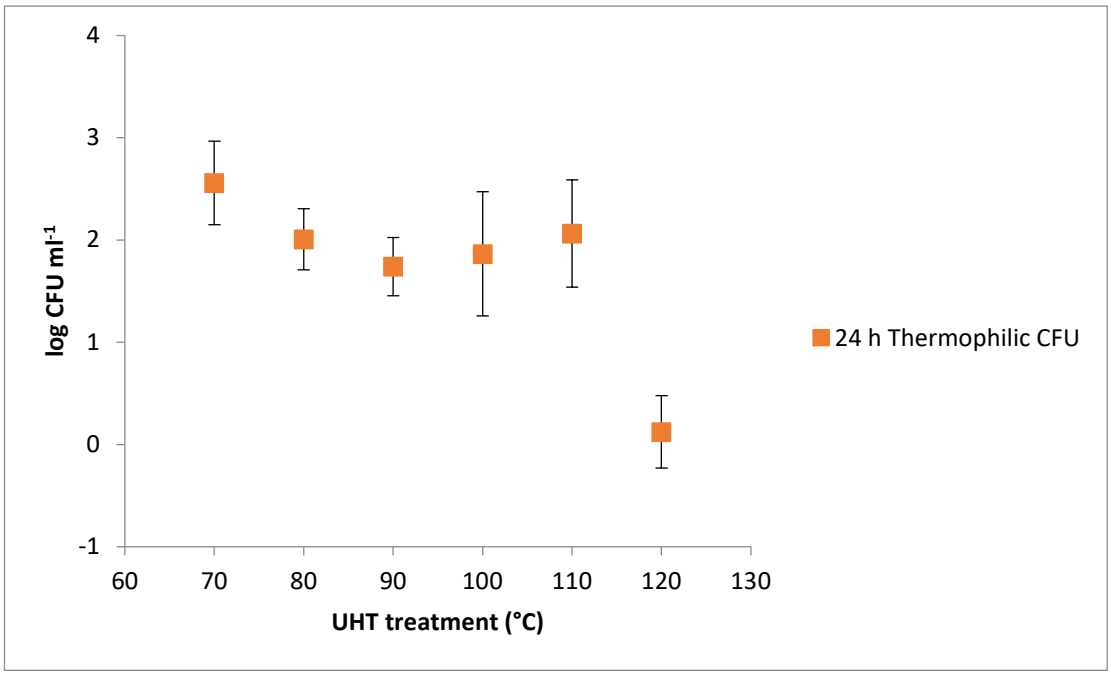
Watterson, M. J., Kent, D. J., Boor, K. J., Wiedmann, M. and Martin, N. H. (2014) 'Evaluation of dairy powder products implicates thermophilic sporeformers as the primary organisms of interest', *J Dairy Sci,* 97(4), pp. 2487-97. DOI: 10.3168/jds.2013-7363

Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis.* Springer Publishing Company, Incorporated. DOI: 10.1007/978-0-387-98141-3

Zhang, Z., Liu, W., Xu, H., Aguilar, Z. P., Shah, N. P. and Wei, H. (2015) 'Propidium monoazide combined with real-time PCR for selective detection of viable *Staphylococcus aureus* in milk powder and meat products', *J Dairy Sci,* 98(3), pp. 1625-33. DOI: 10.3168/jds.2014-8938

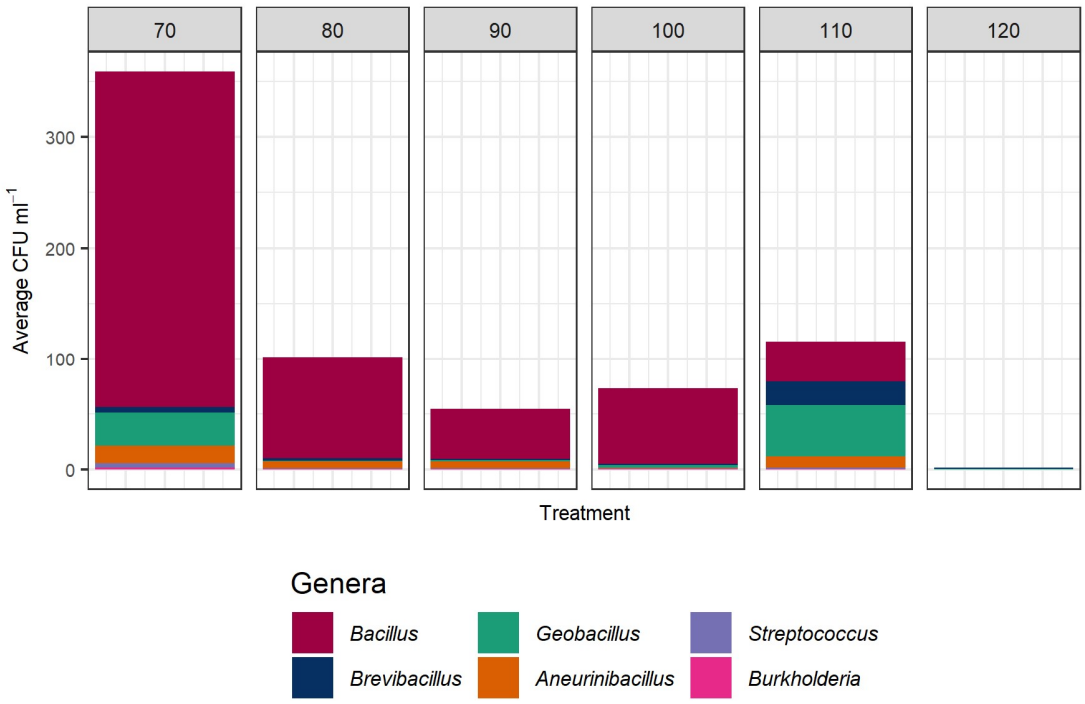
Zou, M. and Liu, D. (2018) 'A systematic characterization of the distribution, biofilm-forming potential and the resistance of the biofilms to the CIP processes of the bacteria in a milk powder processing factory', *Food Res Int,* 113, pp. 316-326. DOI: 10.1016/j.foodres.2018.07.020

2.8 Supplemental Figures



Supplemental Figure 2.1. HHR thermophilic spore-forming log CFU ml^{-1}.

Mean and standard deviation (SD) log CFU ml^{-1} for each UHT treatment following enrichment for thermophilic bacteria at 55°C x 24 h post UHT treatment. Samples selected for HHR sporeformers were pour plated in PCSMA, incubated at 55°C for 48 h before counting CFU.



Supplemental Figure 2.2. Mean CFU of genera level taxonomic classification of thermophilic sporeformers per UHT treatment.

Mean CFU of average genera taxonomic classification per UHT treatment employed. Represents a combination of Figure 2.1A and Supplemental Figure 2.1.

Chapter 3. Mesophilic sporeformers identified in whey powder by using shotgun metagenomic sequencing

Included as published in Applied and Environmental Microbiology

(doi: 10.1128/AEM.01305-18)

Authors: Aoife J. McHugh, Conor Feehily, John T. Tobin, Mark A. Fenelon, Colin Hill, Paul D. Cotter

Contributions: Candidate performed heat treatments, culture-based analysis, DNA extraction, sequencing library preparation, bioinformatic and statistical analysis.

Candidate drafted and edited chapter

JTT supplied powder samples.

All authors contributed, revised and approved manuscript.

3.1 Abstract

Spoilage and pathogenic spore-forming bacteria are a major cause of concern for producers of dairy products. Traditional agar-based detection methods employed by the dairy industry have limitations with respect to their sensitivity and specificity. The aim of this study was to identify low-abundance sporeformers in samples of a powdered dairy product, whey powder, produced monthly over one year, using novel culture-independent shotgun metagenomics-based approaches. Although mesophilic sporeformers were the main target of this study, in one instance thermophilic sporeformers were also targeted using this culture-independent approach. For comparative purposes, mesophilic and thermophilic sporeformers were also tested for within the same sample using culture-based approaches. Ultimately, the approaches taken highlighted differences in the taxa identified due to treatment and isolation methods. Despite this, low levels of transient, mesophilic, and in some cases potentially pathogenic, sporeformers were consistently detected in powder samples. Although the specific sporeformers changed from one month to next, it was apparent that 3 groups of mesophilic sporeformers, namely, *Bacillus cereus*, *Bacillus licheniformis*/*Bacillus paralicheniformis*, and a third, more heterogeneous group containing *Brevibacillus brevis,* dominated across the 12 samples. Total thermophilic sporeformer taxonomy was considerably different to mesophilic taxonomy, as well as the culturable thermophilic taxonomy in the one sample analysed by all four approaches. Ultimately, through the application of shotgun metagenomic sequencing to dairy

powders, the potential for this technology to facilitate the detection of undesirable bacteria present in these food ingredients is highlighted.

3.1.1 Importance

The ability of sporeformers to remain dormant in a desiccated state is of concern from a safety and spoilage perspective in dairy powder. Traditional culturing techniques are slow and provide little information without further investigation. We describe the identification of mesophilic sporeformers present in powders produced over one year, using novel shotgun metagenomic sequencing. This method allows detection and identification of possible pathogens and spoilage bacteria in parallel. Strain-level analysis and functional gene analysis, such as identification of toxin genes, were also performed. This approach has the potential to be of great value with respect to the detection of spore-forming bacteria and could allow a processor to make an informed decision surrounding process changes to reduce the risk of spore contamination.

3.2 Introduction

Milk and resultant dairy products can become contaminated by bacteria from a number of sources, including production and processing facility contaminants. Soil, bedding, feed, feces and teat surface all harbor bacteria that can transfer to raw milk (Gleeson et al., 2013, Doyle et al., 2017), with milking and housing practices are a potential contributor to raw milk contamination (Miller et al., 2015a). Although many of the bacteria present in milk are killed by traditional processing techniques, bacterial spores can survive heat treatment and desiccation (Gleeson et al., 2013, Setlow, 2003). Furthermore, within processing facilities, microbial biofilms formed on equipment surfaces can be persistent. These are frequently resistant to cleaning and cells, including spore-forming bacteria, from these biofilms can slough off to contaminate products (Pérez-Rodríguez et al., 2008, Faille et al., 2014). Furthermore, spores, regardless of their origin, can withstand further processing and remain in a dormant form in powdered dairy products thereafter (Doyle et al., 2015, Gopal et al., 2015).

On the basis of culture-based analyses, sporeformers identified frequently in dairy powders include representatives of the genera *Bacillus*, *Geobacillus*, *Anoxybacillus*, *Lysinibacillus*, *Brevibacillus*, *Paenibacillus* (Ronimus et al., 2003, Sadiq et al., 2016) as well as *Clostridium* (Barash et al., 2010, Buehner et al., 2015). Some of these sporeformers, including *Bacillus licheniformis*, *Bacillus subtilis*, *Bacillus amyloliquefaciens*, members of *Bacillus cereus sensu lato* (De Jonghe et al., 2010) and *Geobacillus* species (Seale et al., 2012), are associated with spoilage of dairy products. Also, some members of *B. cereus sensu lato* have the potential to cause

food poisoning. More specifically, *B*. *cereus* can cause diarrheal or emetic food poisoning. Diarrheal food poisoning is caused by enterotoxins, i.e., nonhemolytic enterotoxin (Nhe), cytotoxin K (CytK) or hemolysin BL (HBL), with symptom onset occurring 8 to 16 h after ingestion (Granum and Lund, 1997, Ehling-Schulz et al., 2006). Emetic food poisoning is caused by emetic toxin cereulide (Ces), which is produced by a nonribsomal peptide synthetase and results in vomiting within a few hours of ingestion (Ehling-Schulz et al., 2006). Toxin-producing *Clostridium* species, such as *Clostridium botulinum* and *Clostridium perfringens* (Doyle et al., 2015), are also a concern but are less common in dairy powder than bacteria of the class Bacilli (Miller et al., 2015b).

Despite safety and economic concerns, there is little legislation governing the total numbers of permissible sporeformers in dairy powders. However, processors and customers generally set strict guidelines to ensure high standards. U.S. powders destined for international customers have limits of < 500 CFU g^{-1} thermophilic sporeformers and < 1,000 CFU g^{-1} mesophilic sporeformers (Watterson et al., 2014). These limits reflect the fact that traditional culture-based detection and enumeration methods for spore-forming bacteria rely on a variety of temperature treatments to differentiate between heat-resistant and highly heat-resistant sporeformers, incubation temperatures (to differentiate among psychrotolerant, mesophilic and thermophilic sporeformers), and incubation conditions (to differentiate between aerobic and anaerobic sporeformers) (Watterson et al., 2014, Miller et al., 2015b). A number of selective and chromogenic agars may also be employed to select for and/or identify pathogenic groups, such as *Clostridium*

species and *B*. *cereus* group (Weenk et al., 1995, Tallent et al., 2012). Further biochemical or molecular methods can be subsequently utilized to identify the species isolated (Gleeson et al., 2013).

The heavy reliance on culture-based assays has been highlighted as a concern in recent years. These methods are labor intensive, requiring many media types, sample treatments, incubation temperatures, and conditions per sample to get an overview of total numbers of different groups of bacteria present. These need to be followed by additional testing to identify or confirm identity of the species present and, where necessary, further analysis to determine if species present contain toxin genes (Wehrle et al., 2009).

Here, shotgun metagenomic DNA sequencing (Sharpton, 2014) is employed to identify bacteria present in dairy powders, as an extension of the recent application of the technology to identify pathogenic strains in a fermented dairy beverage (Walsh et al., 2017) or spoilage microorganisms in cheese (Quigley et al., 2016). Shotgun metagenomics can generate a considerable volume of data, enabling the detection of culturable and nonculturable microorganisms, with accurate identification to species level and sometimes strain level. It also allows an investigation of specific gene sequences of interest, such as, for example, antimicrobial resistance or toxin-encoding genes (Bengtsson-Palme et al., 2014, Leonard et al., 2015). However, this analysis remains expensive and poses data analysis-related difficulties due to the frequently large quantity of data generated (Sharpton, 2014). With a view to beginning to bridge the gap toward the application of this powerful technology to food quality and safety testing, this study employed

shotgun metagenomics to test 12 powdered dairy samples produced monthly at the same location over one calendar year to determine the mesophilic spore content of these powders. For the purpose of comparison, one sample (from August [A]) was examined in greater depth to assess the mesophilic spore content (A.M), thermophilic spore content (A.T), as well as the culturable mesophilic spore content (A.M.P), and culturable thermophilic spore content (A.T.P).

3.3 Materials and methods

3.3.1 Sample preparation and enrichment

Twelve (i.e. one per month for a calendar year) 5 g whey powder samples from a single supplier were suspended aseptically at a concentration of 10% (wt/vol) in sterile ¼ strength Ringer's solution (Sigma-Aldrich). Each reconstituted sample was incubated at 80°C for 12 min to select for spore-forming bacteria as previously described (Miller et al., 2015a, Miller et al., 2015b). An aliquot was then plated onto Brain Heart Infusion (BHI) agar before incubation at mesophilic (30°C) and thermophilic (55°C) suitable temperatures and CFU g^{-1} values were calculated.

Following spore selection, in which 10% (wt/vol) reconstituted samples were incubated at 80°C for 12 min, reconstituted samples were enriched in a manner consistent with that previously employed to select for low abundance mesophilic sporeformers, but with incubation at 30°C for 48 h as opposed to 32°C as previously documented (Miller et al., 2015a, Miller et al., 2015b). This is shown in Supplemental Figure 3.1.

To facilitate an investigation into the extent to which plating altered the identity of the populations of sporeformers detected, powder sourced in August [A], was treated as described above but with incubation at 55°C to select for thermophilic sporeformers (A.T). Both of the enriched August samples were plated onto BHI agar, before incubating at the corresponding mesophilic (30°C; A.M.P) and thermophilic (55°C; A.T.P) temperatures for 48 h (Supplemental Figure 3.1). DNA was extracted from the surface of these agar plates as described below.

3.3.2 DNA extraction and library preparation

A total of fifty millilitres of samples that were reconstituted, heat-treated, and enriched by incubation for 48 h at 30°C were centrifuged at 900 x g for 20 min. The resultant pellet was discarded, and supernatant centrifuged at 4500 x g for 20 min. Pellets were washed in 1 ml sterile ¼ Ringer's solution and centrifuged at 13,000 x g for 2 min. Washing was repeated twice more, centrifuging for 1 min each time at 13,000 x g. DNA was extracted from the resultant pellet as described below. For DNA extraction from a combination of all colonies on the surface of agar plates, 5 ml ¼ Ringer's solution was added and colonies were washed off using a sterile Lazy-L spreader (Sigma-Aldrich). An aliquot of 4 ml of the resultant fluid was removed and centrifuged at 4500 x g for 20 min. The resultant pellet was suspended in 1 ml ¼ Ringer's solution and centrifuged at 13000 x g for 2 min. The pellet was washed twice more in 1 ml of Ringer's solution, and centrifuged at 13000 x g. Before the final centrifugation 200 µl was removed from the mix to a sterile tube. This was centrifuged at 13000 x g for 1 min. This pellet corresponded to 0.8 ml of culture from original 4 ml.

All pellets were resuspended in 150 µl of 45 mg ml^{-1} lysozyme and incubated at 37°C for 30 min. Samples were centrifuged at 13000 x g for 1 min, and the supernatant was removed. From this point, the MoBio PowerFood DNA Isolation kit protocol was followed, including the use of the alternative lysis step for difficult-to-lyse cells. DNA was eluted in 60 µl of 10mM Tris-HCl and stored at -20°C. DNA was quantified and quality checked using the Qubit double-stranded DNA (dsDNA) high

sensitivity assay kit (Bio-Sciences, Dublin, Ireland) and by visualising on a 1% (wt/vol) agarose gel.

Samples were prepared for metagenomic sequencing according to Illumina Nextera XT Library preparation kit guidelines and sequenced on an Illumina Miseq sequencing platform at the Teagasc DNA Sequencing Facility with a 2 x 250 V2 kit using standard Illumina sequencing protocols.

3.3.3 Bioinformatics pipeline

Raw metagenomic shotgun reads were checked for the presence of human and bovine reads filtered on the presence of quality and quantity, and trimmed to 170 bp with a combination of Picard tools (http://broadinstitute.github.io/picard/) and SAMtools (Li et al., 2009). Kraken with a filter threshold of 0.2 (Wood and Salzberg, 2014) and SUPER-FOCUS (Silva et al., 2016) were used to determine microbial composition to species level and biological functions, respectively. MetaPhlAn2 (Truong et al., 2015) and Kaiju (Menzel et al., 2016), were also utilized to determine microbial composition, for comparison to Kraken results. Eleven *B*. *cereus* toxin-associated genes (Table 3.1) were downloaded from the NCBI database and aligned to sequence reads using Bowtie2 (Langmead and Salzberg, 2012).

Table 3.1. Toxin genes aligned to sequence reads.

Toxin	Toxin gene	Gene identifier	Genome	Description
Cytotoxin K	*cytK*	gi|30018278:c1092459-1091449	*Bacillus cereus* ATCC 14579 chromosome	BC1110 (cytotoxin K)
Nonhemolytic enterotoxin	*nhe*	gi|30018278:1765248-1766408	*Bacillus cereus* ATCC 14579 chromosome	BC1809 (nonhemoyltic enterotoxin lytic component L2)
Nonhemolytic enterotoxin	*nhe*	gi|30018278:1766446-1767654	*Bacillus cereus* ATCC 14579 chromosome	BC1810 (nonhemolytic enterotoxin lytic component L1)
Hemolysin BL	*hbl*	gi|30018278:c3062258-3060858	*Bacillus cereus* ATCC 14579 chromosome	BC3101 (hemolysin BL binding component precursor)
Hemolysin BL	*hbl*	gi|30018278:c3063761-3062634	*Bacillus cereus* ATCC 14579 chromosome	BC3102 (hemolysin BL binding component precursor)
Hemolysin BL	*hbl*	gi|30018278:c3065018-3063798	*Bacillus cereus* ATCC 14579 chromosome	BC3103 (hemolysin BL lytic component L1)
Hemolysin BL	*hbl*	gi|30018278:c3066399-3065080	*Bacillus cereus* ATCC 14579 chromosome	BC3104 (hemolysin BL lytic component L2)
Cereulide	*cesA*	gi|190015498:c35141-25017	*Bacillus cereus* strain AH187 plasmid pCER270	CesA (cereulide synthetase A)
Cereulide	*cesB*	gi|190015498:c25003-16958	*Bacillus cereus* strain AH187 plasmid pCER270	CesB (cereulide synthethase B)
Cereulide	*cesC*	gi|190015498:c16813-15917	*Bacillus cereus* strain AH187 plasmid pCER270	CesC (ABC transporter ATP binding protein)
Cereulide	*cesD*	gi|190015498:c15900-15094	*Bacillus cereus* strain AH187 plasmid pCER270	CesD (putative permease)

Metagenomes were assembled into contigs using IDBA-UD (Peng et al., 2012) and toxin genes were aligned to these using Bowtie2 (Langmead and Salzberg, 2012). Toxin genes were also aligned to contigs using Mauve version 20150226 (Darling et al., 2004), more specifically progressiveMauve (Darling et al., 2010), using default parameters (default seed weight, determine local collinear blocks [LCBs], full alignment, iterative refinement, min LCB weight = default, sum-of-pairs LCB scoring). PanPhlAn (Scholz et al., 2016) was utilized to determine *B*. *cereus* strain profiles in each sample. A total of thirty two complete *B. cereus* genome sequences were downloaded from the NCBI database and a pangenome was generated and compared to all samples. PanPhlAn outputs were visualized using GraPhlAn (Asnicar et al., 2015). Spearman correlations with Benjamini-Hochberg corrections for multiple comparisons were made in R using package Hmisc and visualized using corrplot. Diversity was analysed using the R vegan package and data visualization was performed using ggplot2. Sequence reads were aligned to Bagel 3 bacteriocin database (van Heel et al., 2013) using DIAMOND (Buchfink et al., 2015), to determine the bacteriocin potential of the bacteria identified.

3.4 Results

3.4.1 Shotgun metagenomics can be used to identify and determine the functional potential of low-abundance sporeformers present in dairy powders.

A total of 12 whey powder samples, produced monthly at the same location over one calendar year, were sampled upon exit from the spray drier prior to packing and collected for spore analysis. Following spore-pasteurization of reconstituted powders, in which samples were incubated at 80°C for 12 min to select for spore-forming bacteria, a total of < 400 CFU g^{-1} mesophilic and thermophilic sporeformers were enumerated by plating whey solutions on brain heart infusion (BHI) agar and incubating at 30˚C and 55˚C, respectively, for 24 h. These are within previously described consumer specified limits of < 500 CFU g^{-1} thermophilic sporeformers and < 1,000 CFU g^{-1} mesophilic sporeformers (Watterson et al., 2014).

To facilitate a more in-depth characterization, sporeformer enrichment was performed and DNA extraction and shotgun metagenomic sequencing was completed. The enrichment approach taken addressed the risk of reads arising from sequencing of DNA from dead cells, but had a drawback of under-representing spores which did not germinate under the conditions employed. The average number of reads after quality filtering and trimming per sample was 1,106,747. Kaiju, Kraken and MetaPhlan2 were all used to assess each package's relative ability to taxonomically assign reads, and results from all three tools are presented. Kraken assigned the greatest percentage of reads at the species level. This coupled with the previously reported possibility of high levels of false positives resulting from

Kaiju assignment (Piro et al., 2017) and the fact MetaPhlAn2 works off only a subset of marker genes per species (Truong et al., 2015), lead to Kraken being employed preferentially. Kraken was applied with a filter threshold of 0.2 to increase precision without detrimentally impacting sensitivity. Furthermore, to reduce the possibility of false positives (Piro et al., 2017), taxa were included only if present at a minimum of 1% relative abundance in at least one sample, otherwise reads were assigned as "others".

B. *cereus* was found to be the dominant species in 7 of the 12 monthly mesophilic sporeformer enriched samples, i.e., those from January, February, March, May, July, October and November. *Bacillus thuringiensis*, a member of *B*. *cereus sensu lato*, was the next most abundant species in these samples (Figure 3.1). Among the other mesophilic sporeformer-enriched samples, *B*. *licheniformis* was present as the dominant species in April and August. It was also present in the September and December samples but not dominant. *B*. *subtilis* was dominant in the September sample but also detected at lower relative abundance in August and December. *B*. *paralicheniformis* was the dominant species in the December sample while also present, but not dominant in April, August and September (Figure 3.1). The June sample was dominated by *Brevibacillus brevis* and *Streptococcus thermophilus*. *S*. *thermophilus* was also detected in lower levels in other samples too, despite not being a sporeformer. All species identified as dominant were confirmed by Kaiju- and MetaPhlAn2-derived results (Supplemental Figure 3.2 & Supplemental Figure 3.3).

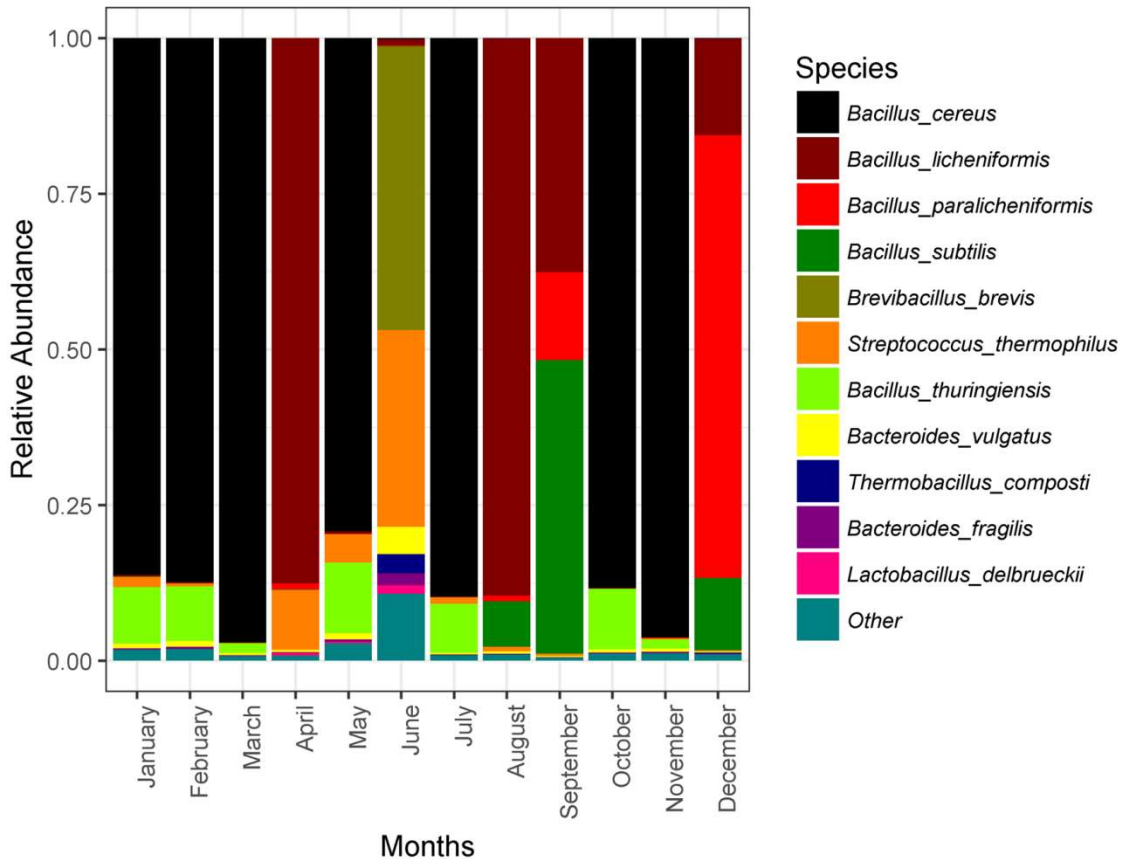


Figure 3.1. Relative abundance of mesophilic spore-forming enriched species present per sample.

Relative abundance of the most abundant species (shotgun metagenomic reads assigned by Kraken with 0.2 filter threshold and minimum of 1% relative abundance in at least one sample) in samples enriched for mesophilic sporeformers.

Spearman correlations were utilized to evaluate the population relationships in samples with taxonomic analysis and functional gene analysis. Spearman correlations are used for nonnormally distributed data that is either skewed or ordinal. Statistical analysis was performed on correlations, asymptotic P values were generated using rcorr in the R package Hmisc and corrected with Benjamini-Hochberg corrections for multiple comparisons. Significant negative correlation ($p < 0.05$) between *B*. *cereus* and *B*. *licheniformis* and between *B*. *cereus* and *B*. *paralicheniformis* were noted. Significant positive correlations ($p < 0.05$) were seen between *B*. *licheniformis* and *B*. *paralicheniformis*, *Bacteriodes vulgatus* and *Bacteriodes fragilis*, *B*. *vulgatus* and others, and *B*. *fragilis* and others. As expected, many sporulation-associated functional gene groups were significantly positively correlated ($p < 0.05$), including those for dormancy and sporulation with those for virulence disease/defense, regulation/cell signalling, iron acquisition/metabolism, and cell wall/capsule formation (Supplemental Figure 3.4).

Fasta reads were aligned to Bagel 3 bacteriocin database using DIAMOND to determine if bacteriocin production potential could be influencing current and future relationship dynamics. This showed a number of potential type I, type II and type III bacteriocin genes in each sample, with highest relative abundance per sample going to type I bacteriocin genes (Figure 3.2). Bacteriocin genes of note included two lichenicidin peptide genes present in all samples containing *B*. *licheniformis*, genes for sporulation killing factor skfA and subtilosin are seen in samples containing *B*. *subtilis*, and thuricin genes in samples containing *B*. *thuringiensis* (Figure 3.2A).

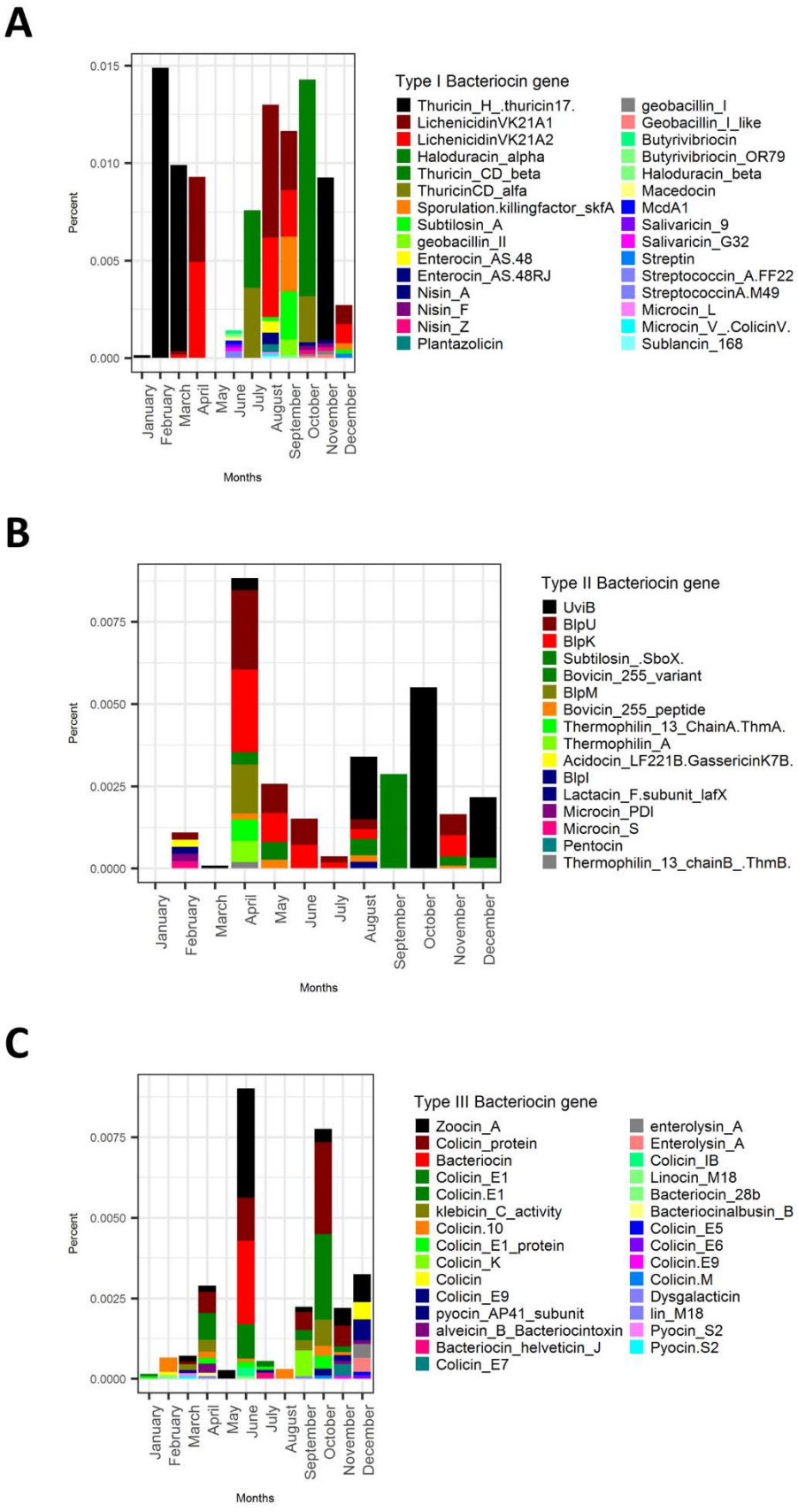


Figure 3.2. Percentage of total reads per sample attributed to bacteriocin genes found in Bagel 3 database.

[Figure 3.2 continued]

A. Percentage of reads attributed to type I bacteriocin genes.

B. Percentage of reads attributed to type II bacteriocin genes.

C. Percentage reads attributed to type III bacteriocin genes.

The August sample was selected for a parallel culture-based analysis and was enriched for culturable mesophilic sporeformers by culturing on BHI agar and incubating at 30°C for 48 h, following 48 h enrichment of a spore-pasteurised sample at 30°C. This was labeled August plate-cultured mesophilic sporeformer enriched sample (A.M.P). An additional aliquot of the same spore-pasteurised sample was enriched for thermophilic sporeformers, by incubating at 55°C for 48 h; this was labeled August thermophilic sporeformer enriched sample (A.T). Subsequently this sample was plated on BHI agar and incubated at 55°C for 48 h. This was labeled as August plate-cultured thermophilic sporeformer-enriched sample (A.T.P). The original August mesophilic sporeformer enriched sample was relabeled A.M for the purpose of comparison. In the case of the plate cultured samples (A.M.P and A.T.P), colonies were scraped off, and DNA extracted in all instances and sequenced as previously described. Comparative analysis revealed that the A.M.P and A.M samples had a very similar profile, highlighting that plating did not bias results (Figure 3.3). However, the taxonomic profile of the A.T sample and A.T.P samples were very different, i.e., the dominant species present in the A.T sample was *Thermoanaerobacterium thermosaccharolyticum*, whereas the equivalent post-plating sample (A.T.P) was comprised of *B*. *brevis*, *Bacteroides* species as well as a variety of other species at low abundance (Figure 3.3).

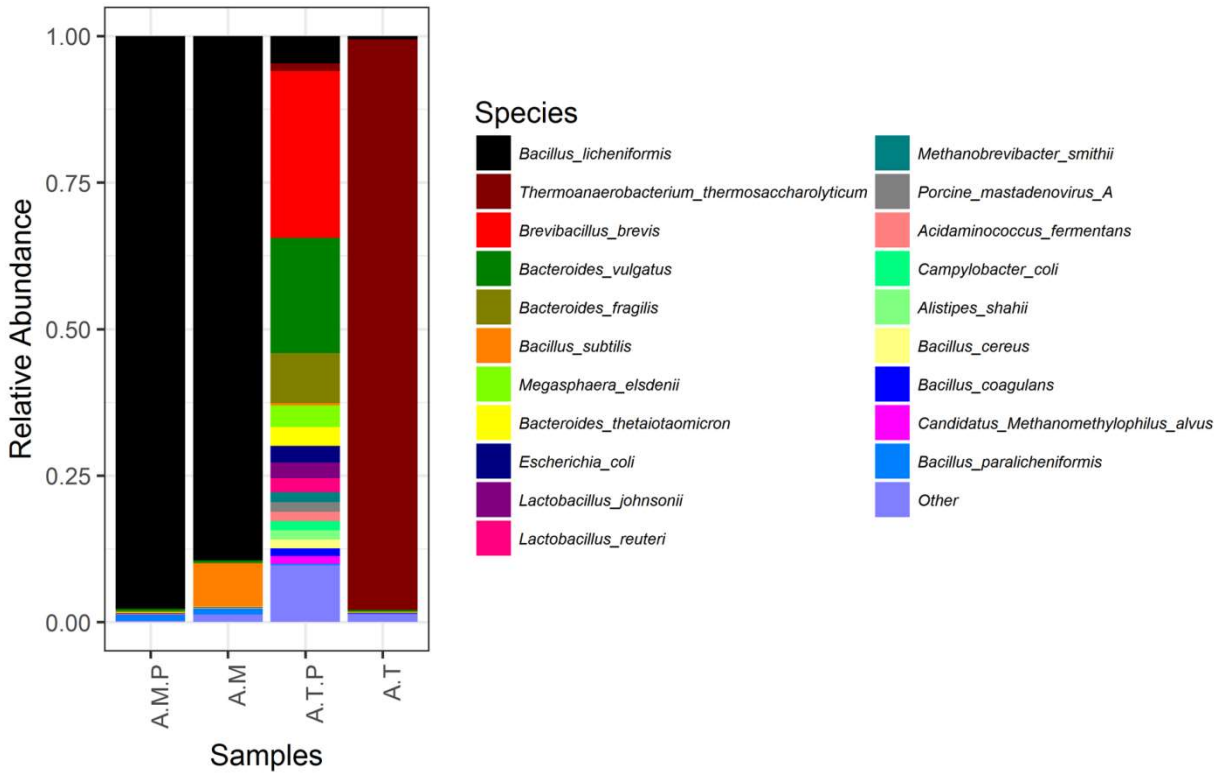


Figure 3.3. Relative abundance of spore-forming enriched species in August sample.

Relative abundance of the most abundant species (assigned by Kraken with a 0.2 filter threshold and with minimum 1% relative abundance in at least one sample) in the August sample subject to different enrichment temperatures and conditions.

3.4.2 Beta diversity analysis highlights that samples cluster according to the dominant species present

Bray-Curtis beta diversity analysis of mesophilic sporeformer reads showed that the 12 samples clustered into 3 distinct groups (individual data points within each group being co-located; Figure 3.4A). One cluster consisted of samples that contained *B*. *cereus* (i.e., January, February, March, May, July, October, November); a second cluster consisting of samples from months April, August, September and December, which all contain high relative abundance of *B. licheniformis*; and a third corresponded to the June sample, which had high relative abundance of *B*. *brevis* (Figure 3.4A).

Alpha diversity analysis did not reveal any notable pattern other than the observation that the June, September and December samples had the highest diversities (Figure 3.4B). As might be expected from taxonomic results, beta diversity analysis of the August sample, which was enriched in a variety of ways, showed that A.M and A.M.P samples were least dissimilar, whereas the A.T and A.T.P samples differed from each other and from A.M and A.M.P (Figure 3.4C). Among these samples, alpha diversity in A.T.P was higher than all others (Figure 3.4D).

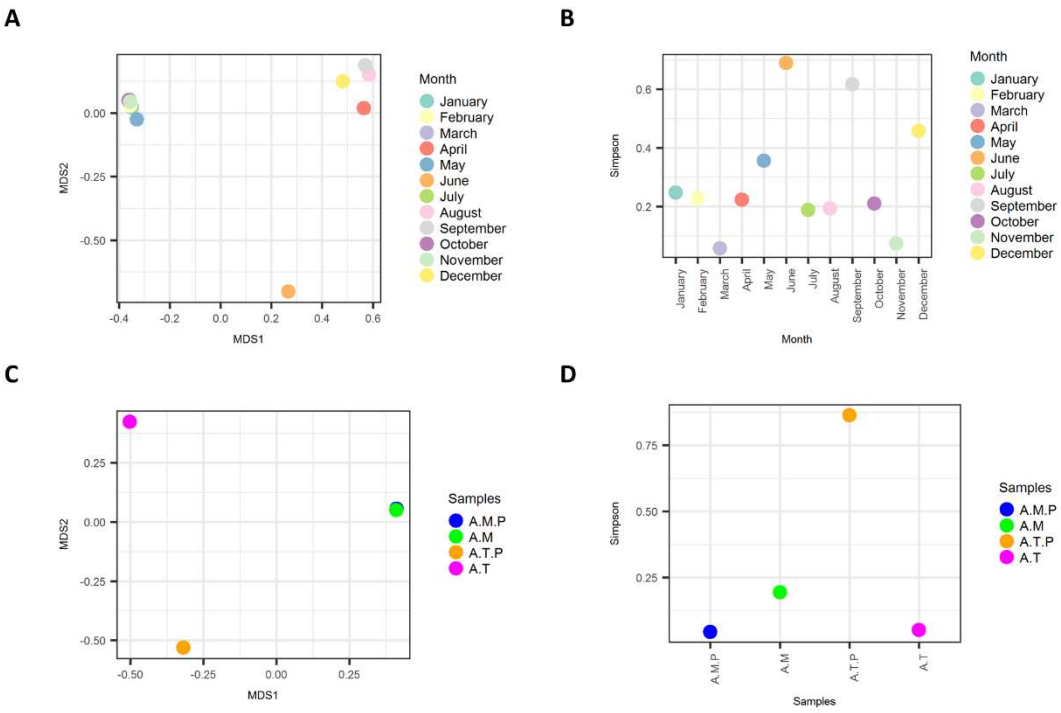


Figure 3.4. Alpha and beta diversity metrics for mesophilic sporeformer-enriched samples and August samples subject to different incubation temperatures and conditions.

A. Bray-Curtis weighted classical multidimensional scaling (MDS) of mesophilic sporeformer-enriched samples.

B. Simpson alpha diversity of mesophilic sporeformer-enriched samples.

C. Bray-Curtis weighted classical multidimensional scaling (MDS) of August samples.

D. Simpson alpha diversity of August samples in different conditions.

3.4.3 Toxin gene analysis revealed presence of potentially toxigenic *B. cereus*

Further analyses was performed to determine if the *B. cereus* strains detected in some samples might be capable of causing emetic and enterotoxic food poisoning. For this purpose, toxin genes were aligned to sequence reads and verified by aligning them to assembled contigs. The genes used and the genome sequences from which they originated are shown in Table 3.1. Of the 11 toxin genes screened for among all 12 mesophilic enriched samples, the two nonhemolytic enterotoxin (Nhe) genes; *nhe* L1 and *nhe* L2*,* were detected in all 7 samples previously shown to contain high relative abundances of *B. cereus*. The cytotoxin K-encoding gene was detected in 3 samples (from March, October and November). The presence of toxin genes in the sequence reads was verified by alignment and visualization using MAUVE. Toxin gene sequences were aligned to sample contigs, which were assembled using IDBA-UD (Supplemental Figure 3.5). Among samples in which all 3 toxin genes were present, toxin genes accounted for close to 0.1% of reads and, when just the two *nhe* genes were present, they accounted for close to 0.06% of reads (Figure 3.5).

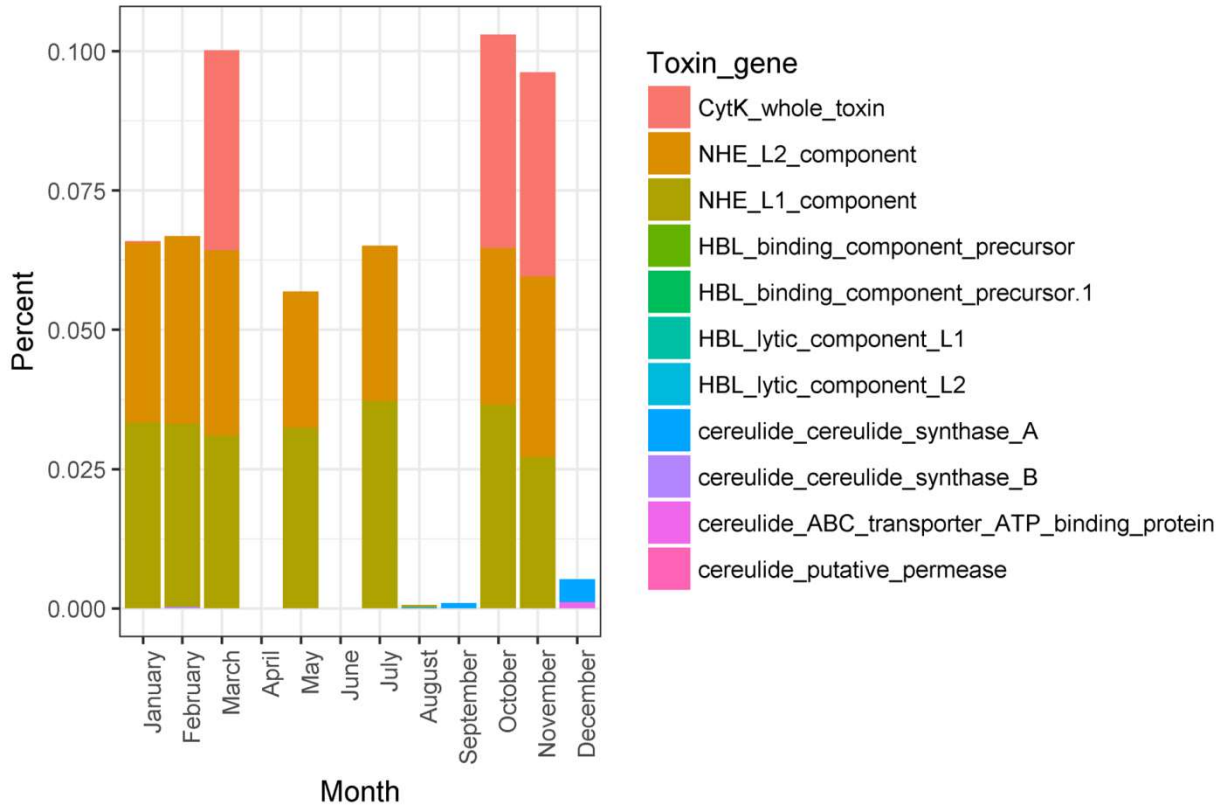


Figure 3.5. Percentage of total reads attributed to *B. cereus* toxin genes.

3.4.4 Strain level analysis revealed the absence of evidence for persistent contamination

As the samples that contained high relative abundances of *B*. *cereus* appeared to have different toxin profiles, it was decided to carry out PanPhlAn-based strain analysis to more accurately determine if one strain was dominating across these samples, potentially indicating strain persistence in the processing facility, or transient colonization by different strains. Sequence reads for all samples were aligned to a pangenome created from 32 complete *B. cereus* genomes downloaded from the NCBI database. This analysis established that the presence/absence of *B*. *cereus* was consistent with previous Kraken analysis. PanPhlAn showed the 7 samples containing *B*. *cereus* contained reads that clustered with 5 strains of *B*. *cereus* by Euclidean distance, as visualized using GraPhlAn (Figure 3.6). The percentage match from sequence reads to the 32 genomes used in the pangenome was also examined (Table 3.2). Overall, as shown in Table 3.2, none of the strains identified in the 7 samples are a complete match to any of the 32 strains with which they were compared. Similarly, none of the strain identified in the 7 samples appear to be exactly the same, with January and February samples being most alike (Figure 3.6).

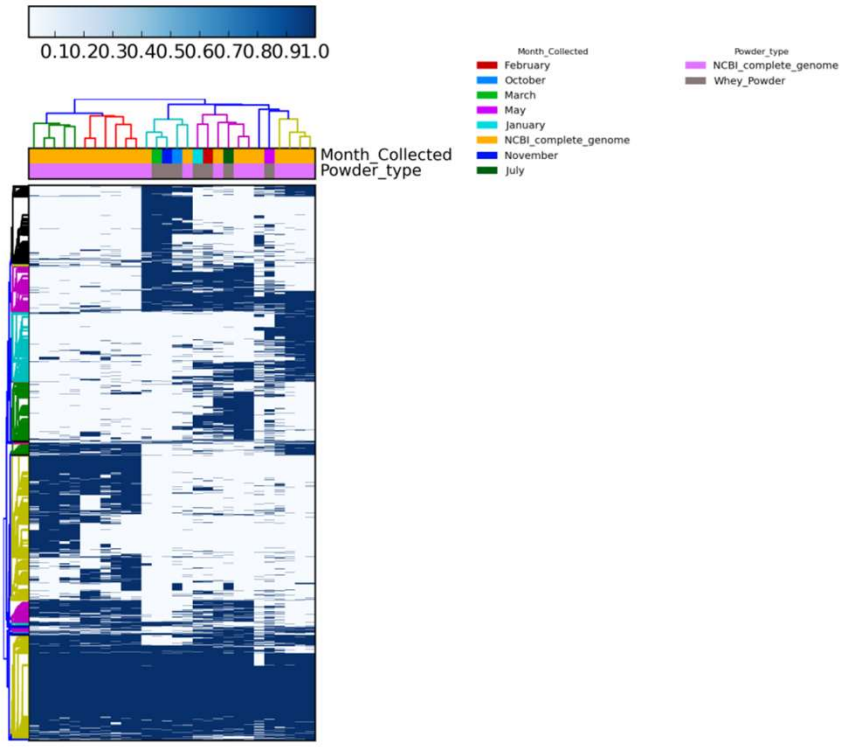


Figure 3.6. GraPhlAn visualization of PanPhlAn clustering of *B*. *cereus* strains.

GraPhlAn visualization of PanPhlAn clustering of 7 samples containing *B. cereus* with representative strains of pangenome. The 7 samples contain reads that cluster with at least 5 distinct strains.

Table 3.2. Percentage match of *B. cereus* strains present in samples to 32 reference strains.

	Percentage match to thee 32 genomes used in the pangenome							
ENA assembly no.	No. of genes	January	February	March	May	July	October	November
GCA_000007825	5554	14.3	13.9	13.3	15.9	14.4	13.4	13
GCA_000008005	5338	52	53.3	89.4	53	52.4	69.2	85.8
GCA_000011625	5868	46.9	47.5	35.5	42.4	44.5	38.5	34.9
GCA_000013065	5521	67.6	68.6	49.9	59.2	76.7	51.7	48.8
GCA_000021205	5366	14.8	14.7	14.4	16.6	15.4	14.5	14.1
GCA_000021225	5637	65.8	66.1	50.1	59.6	79.3	51.7	49.1
GCA_000021305	5749	12.8	12.5	12.4	14.1	13.2	12.5	12.2
GCA_000021785	5595	47.2	47.4	36.2	43.7	43.5	38.3	35.7
GCA_000022505	5515	46	46.5	37.2	42	43.2	39.5	36.6
GCA_000143605	5586	47.5	48.3	36.4	43.4	44.5	38.7	35.8
GCA_000239195	5374	46.7	46.9	36.8	42.1	43.9	39.6	36.2
GCA_000283675	5575	65.9	66.7	50.2	60.1	79.9	51.2	49.1
GCA_000292415	5427	53.8	55.5	69	53.4	54.5	78.5	67.2
GCA_000635895	5625	13.8	13.9	13.4	15.3	14.2	13.6	13.1
GCA_000724585	5583	37.2	37	40.7	43.5	37	36.5	39.4
GCA_000789315	5781	45.5	45.9	41.5	54.3	47.2	42.7	40.6
GCA_000832385	5721	45	47.4	37	41.7	42.1	38.7	36.4
GCA_000832405	5522	45.8	46.3	36.9	41.7	43	39.1	36.2
GCA_000832525	5653	47	47.2	42.7	55.9	48.5	44.2	41.7
GCA_000832765	5708	46.6	46.9	36.1	43	46.8	41.4	35.5
GCA_000832805	5708	46.4	47	42.1	55.4	48.1	43.6	41.1
GCA_000832845	5332	51.7	52.1	44.3	61.4	53	46.4	43.3
GCA_000832865	6126	41.5	42.4	33.8	38.3	39.1	35.6	33.2
GCA_000833045	5885	46.6	47.2	35.3	42.1	44.3	38.2	34.7
GCA_000835185	5707	46.5	46.8	36.1	43.1	46.7	41.4	35.6
GCA_000978375	5312	14.9	14.9	14.5	16.7	15.2	14.5	14.2
GCA_001277915	5369	14.1	13.9	13.5	15.9	14.2	13.7	13.2
GCA_001518875	5669	14.3	14.4	13.4	16.1	15	13.6	13.2
GCA_001635915	5948	13.1	13.4	12.8	14.6	13.1	12.5	12.5
GCA_001635955	5945	13.2	13.5	12.8	14.6	13.1	12.5	12.6
GCA_001635995	5981	14	14.1	13.2	15.3	14.2	13.2	13.1
GCA_001721145	5362	15.1	15.1	14.5	17.1	15.5	14.6	14.2

3.5 Discussion

Traditionally, the detection and identification of bacterial sporeformers has involved culturing under different temperatures and conditions (Watterson et al., 2014, Miller et al., 2015b), together with the use of selective agars to identify pathogenic species (Weenk et al., 1995, Tallent et al., 2012), followed by further biochemical or molecular approaches for confirmation (Gleeson et al., 2013). This study highlights the ability of next-generation shotgun sequencing of enriched samples to identify low-level persistent or transient spore contamination in a dairy powder. Although many of the species identified are similar to those identified in previous studies using traditional detection methods (Miller et al., 2015b), the approach taken has the potential to rapidly identify these species while allowing strain level analysis and, as a result, the tracking of persistent microbes in products. It highlights that extremely low levels of potentially pathogenic bacteria can be present, although in this instance, these bacteria are unlikely to be from persistent contamination in the processing facility but rather are thought to represent a transient low-level contamination. Due to the sensitivity of this approach, new guidelines and standards may need to be introduced to ensure that the risks associated with detection of low-level contamination of *B*. *cereus* in powders are adequately reflected.

The combined approach of high-temperature treatment coupled with shotgun sequencing employed in this study was selected for a number of reasons. Prior to high-temperature activation, the spore-forming bacteria present in dairy powders are likely to be in a spore form, from which DNA extraction can be difficult

(Wielinga et al., 2011). Furthermore, in the absence of enrichment, the possibly high level of DNA remaining from deceased bacteria and the presence of host bovine cells in the powders could also be an issue, as DNA from these sources will also be sequenced when untargeted shotgun metagenomic sequencing is employed (Sharpton, 2014), thereby necessitating the use of additional steps, and expensive kits and reagents to deplete DNA from these other sources (Rudi et al., 2005, Forghani et al., 2015, Feehery et al., 2013, Yokouchi et al., 2006, Binga et al., 2008). DNA amplification would likely be necessary thereafter due to the low yield of DNA from the low level of sporeformers present in samples. Amplification methods add extra bias, costs, and potential for contamination (Binga et al., 2008, Ahsanuddin et al., 2017, Thoendel et al., 2017). A more targeted, economical approach was used to focus on sporeformers that could potentially germinate in rehydrated powder given favorable temperatures, without the extra bias and cost of kits. It should be noted that low-speed centrifugation utilized in the approach from which solids were discarded prior to DNA extraction could result in loss of some spore-forming bacteria that have a high affinity for denatured milk protein solids.

This targeted, culture-independent approach was compared with corresponding culture-dependent approaches using the August sample as a representative test case. The thermophilic sporeformer enriched August sample (A.T), from which DNA was extracted, was also subject to culturing on BHI agar at thermophilic temperatures (A.T.P) before pooling colonies and extracting DNA. DNA from both A.T and A.T.P samples were sequenced (Supplemental Figure 3.1) and compared.

The results were different in that the dominant species identified following enrichment (i.e., without culturing on BHI agar) was *T. thermosaccharolyticum*, while this species was not detected in the cultured sample, potentially showing bias due to agar and conditions used. *T. thermosaccharolyticum* has previously been detected in dried vegetables (Postollec et al., 2012), and more recently, *Thermoanaerobacterium* spp. has been detected in the core microbiome of raw milk (Rodrigues et al., 2017). *T*. *thermosaccharolyticum* is a thermophilic, anaerobic sporeformer and was previously classified as a member of the *Clostridium* genus, although subsequently reclassified (Collins et al., 1994). It is a known canned food spoilage organism (André et al., 2013) that produces hydrogen and causes swelling in canned foods. This ability to efficiently produce hydrogen makes it a potential important organism for sustainable biohydrogen production (O-Thong et al., 2008, Mtimet et al., 2016). The previous nondetection of the species in dairy powders may relate to an inability to grow on the agar substrates conventionally employed by the dairy industry.

A decision was made to further investigate the potential *B*. *cereus* sequences identified to eliminate the potential that the taxonomic classifier was misassigning other members of *B*. *cereus sensu lato* and *B*. *cereus sensu stricto*, as members of this group are notoriously difficult to differentiate (Liu et al., 2015). Toxin profiling and PanPhlAn analysis confirmed that the strains had the potential to be pathogenic, while also highlighting differences between them. All except one of the samples that contained *B*. *cereus* showed complete alignment with both Nhe toxin-encoding genes. The exception was the February sample, from which one half of

nhe L1 gene was absent. This may be due to misassembly or poor coverage of the genome, or may reflect a natural mutation in this strain (Supplemental Figure 3.5). It should be noted that alignment to toxin genes infers potential ability to produce toxins, but does not conclusively indicate functional toxin presence. Also, a total of 10^5 to 10^9 CFU of toxin producing *B*. *cereus* is needed to cause food poisoning (Granum and Lund, 1997). The highly sensitive, qualitative nature of shotgun metagenomic analysis suggests that further steps will need to be taken to determine the risk associated with the detection of these and other toxin genes in food samples. PanPhlAn analysis showed that the 7 samples containing *B*. *cereus* contained reads that clustered with at least 5 strains of *B*. *cereus* by Euclidean distance suggesting the strains identified in the samples are different and not as a result of persistent contamination in the powder production facility or elsewhere.

The negative correlations between *B*. *cereus* and *B*. *licheniformis*/*B*. *paralicheniformis* raises the possibility of competition to determine the dominant species and the possible inhibition of the other species. The potential bacteriocin genes detected in all samples may impact the relationship dynamics observed and may be having an antagonistic effect on some species currently or in future food products. Lichenicidin is a two-peptide lantibiotic previously shown to be active against pathogenic gram-positive bacteria, including *B*. *cereus* (Begley et al., 2009), and we noted alignment to two peptides associated with lichenicidin in all four samples containing *B*. *licheniformis*. However, potential identification of two peptide genes does not suggest the presence of the whole bacteriocin gene cluster or infer correct post translational modifications to produce a functional bacteriocin.

Positive correlations between virulence, disease, defense, dormancy, and sporulation highlight the need to be cautious of sporeformers in food products. A greater understanding of their relationships will aid the prevention of spoilage and pathogenic species that cause concern in food processing.

Disadvantages that need to be overcome in order to allow for the routine use of the sequencing technologies employed in this study primarily relate to cost of analysis, which is currently too expensive for large-scale routine use. Additionally, there are challenges relating to assembly of genomes from shotgun metagenomic sequencing (Sharpton, 2014) and difficulties arising from insufficient accuracies associated, to different extents, with taxonomic classifiers (Piro et al., 2017). There are some solutions emerging, whereby new lower-cost, rapid sequencers are arriving on the market, with MinION (Brown et al., 2017) leading the way towards quick portable detection systems for microorganisms. Through the generation of more reference genome sequences and good-quality shotgun metagenomic sequencing, reference databases and the accuracy of results will improve.

3.6 Conclusions

This study highlighted monthly diverging contamination patterns in whey powder production, which converged into 3 distinct mesophilic sporeformer population groups from 12 powder samples produced in the same production facility, and has shown that the way in which the powders are treated post production (namely, incubation temperature post reconstitution) influences which bacteria germinate and become dominant. Shotgun metagenomics is a useful tool to delve deeper into the understanding of sporeformers and their relationships in food processing,

although it brings with it its own set of caveats and need for guidelines for use and interpretation of results.

3.7 Declarations

3.7.1 Accession numbers

Sequence data have been deposited in the European Nucleotide Archive (ENA) under the study accession number PRJEB24853.

3.7.2 Acknowledgements including funding source

This Research was funded by Department of Agriculture, Food and the Marine (DAFM) under the FIRM project 'SACCP', reference number 14/F/883. Research in the Cotter lab is also funded by SFI award 'Obesiobiotics' (11/P1/1137). Thanks to Orla O'Sullivan, Fiona Crispie and Laura Finnegan of the Food Bioscience department in Teagasc Moorepark.

3.8 References

Ahsanuddin, S., Afshinnekoo, E., Gandara, J., Hakyemezoglu, M., Bezdan, D., Minot, S., Greenfield, N. and Mason, C. E. (2017) 'Assessment of REPLI-g multiple displacement whole genome amplification (WGA) techniques for metagenomic applications', *J Biomol Tech,* 28(1), pp. 46-55. DOI: 10.7171/jbt.17-2801-008

André, S., Zuber, F. and Remize, F. (2013) 'Thermophilic spore-forming bacteria isolated from spoiled canned food and their heat resistance. Results of a French ten-year survey', *International Journal of Food Microbiology,* 165(2), pp. 134-143. DOI: 10.1016/j.ijfoodmicro.2013.04.019

Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C. and Segata, N. (2015) 'Compact graphical representation of phylogenetic data and metadata with GraPhlAn', *PeerJ,* 3, pp. e1029. DOI: 10.7717/peerj.1029

Barash, J. R., Hsia, J. K. and Arnon, S. S. (2010) 'Presence of soil-dwelling Clostridia in commercial powdered infant formulas', *J Pediatr,* 156(3), pp. 402-8. DOI: 10.1016/j.jpeds.2009.09.072

Begley, M., Cotter, P. D., Hill, C. and Ross, R. P. (2009) 'Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins', *Appl Environ Microbiol,* 75(17), pp. 5451-60. DOI: 10.1128/AEM.00730-09

Bengtsson-Palme, J., Boulund, F., Fick, J., Kristiansson, E. and Larsson, D. G. (2014) 'Shotgun metagenomics reveals a wide array of antibiotic resistance genes and

mobile elements in a polluted lake in India', *Front Microbiol,* 5, pp. 648. DOI: 10.3389/fmicb.2014.00648

Binga, E. K., Lasken, R. S. and Neufeld, J. D. (2008) 'Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology', *ISME J,* 2(3), pp. 233-41. DOI: 10.1038/ismej.2008.10

Brown, B. L., Watson, M., Minot, S. S., Rivera, M. C. and Franklin, R. B. (2017) 'MinION™ nanopore sequencing of environmental metagenomes: a synthetic approach', *GigaScience,* 6(3), pp. 1-10. DOI: 10.1093/gigascience/gix007

Buchfink, B., Xie, C. and Huson, D. H. (2015) 'Fast and sensitive protein alignment using DIAMOND', *Nat Meth,* 12(1), pp. 59-60. DOI: 10.1038/nmeth.3176

Buehner, K. P., Anand, S. and Djira, G. D. (2015) 'Prevalence of thermoduric bacteria and spores in nonfat dry milk powders of Midwest origin', *J Dairy Sci,* 98(5), pp. 2861-6. DOI: 10.3168/jds.2014-8822

Collins, M. D., Lawson, P. A., Willems, A., Cordoba, J. J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H. and Farrow, J. A. (1994) 'The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations', *Int J Syst Bacteriol,* 44(4), pp. 812-26. DOI: 10.1099/00207713-44-4-812

Darling, A. C. E., Mau, B., Blattner, F. R. and Perna, N. T. (2004) 'Mauve: multiple alignment of conserved genomic sequence with rearrangements', *Genome Research,* 14(7), pp. 1394-1403. DOI: 10.1101/gr.2289704

Darling, A. E., Mau, B. and Perna, N. T. (2010) 'progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement', *PLOS ONE,* 5(6), pp. e11147. DOI: 10.1371/journal.pone.0011147

De Jonghe, V., Coorevits, A., De Block, J., Van Coillie, E., Grijspeerdt, K., Herman, L., De Vos, P. and Heyndrickx, M. (2010) 'Toxinogenic and spoilage potential of aerobic spore-formers isolated from raw milk', *International Journal of Food Microbiology,* 136(3), pp. 318-325. DOI: 10.1016/j.ijfoodmicro.2009.11.007

Doyle, C. J., Gleeson, D., Jordan, K., Beresford, T. P., Ross, R. P., Fitzgerald, G. F. and Cotter, P. D. (2015) 'Anaerobic sporeformers and their significance with respect to milk and dairy products', *Int J Food Microbiol,* 197, pp. 77-87. DOI: 10.1016/j.ijfoodmicro.2014.12.022

Doyle, C. J., Gleeson, D., O'Toole, P. W. and Cotter, P. D. (2017) 'Impacts of seasonal housing and teat preparation on raw milk microbiota: a high-throughput sequencing study', *Appl Environ Microbiol,* 83(2). DOI: 10.1128/AEM.02694-16

Ehling-Schulz, M., Guinebretiere, M. H., Monthan, A., Berge, O., Fricker, M. and Svensson, B. (2006) 'Toxin gene profiling of enterotoxic and emetic *Bacillus cereus*', *FEMS Microbiol Lett,* 260(2), pp. 232-40. DOI: 10.1111/j.1574-6968.2006.00320.x

Faille, C., Benezech, T., Midelet-Bourdin, G., Lequette, Y., Clarisse, M., Ronse, G., Ronse, A. and Slomianny, C. (2014) 'Sporulation of *Bacillus* spp. within biofilms: a potential source of contamination in food processing environments', *Food Microbiol,* 40, pp. 64-74. DOI: 10.1016/j.fm.2013.12.004

Feehery, G. R., Yigit, E., Oyola, S. O., Langhorst, B. W., Schmidt, V. T., Stewart, F. J., Dimalanta, E. T., Amaral-Zettler, L. A., Davis, T., Quail, M. A. and Pradhan, S. (2013) 'A method for selectively enriching microbial DNA from contaminating vertebrate host DNA', *PLoS ONE,* 8(10), pp. e76096. DOI: 10.1371/journal.pone.0076096

Forghani, F., Langaee, T., Eskandari, M., Seo, K. H., Chung, M. J. and Oh, D. H. (2015) 'Rapid detection of viable *Bacillus cereus* emetic and enterotoxic strains in food by coupling propidium monoazide and multiplex PCR (PMA-mPCR)', *Food Control,* 55, pp. 151-157. DOI: 10.1016/j.foodcont.2015.02.049

Gleeson, D., O'Connell, A. and Jordan, K. (2013) 'Review of potential sources and control of thermoduric bacteria in bulk-tank milk', *Irish Journal of Agricultural and Food Research,* 52(2), pp. 217-227.

Gopal, N., Hill, C., Ross, P. R., Beresford, T. P., Fenelon, M. A. and Cotter, P. D. (2015) 'The prevalence and control of *Bacillus* and related spore-forming bacteria in the dairy industry', *Front Microbiol,* 6, pp. 1418. DOI: 10.3389/fmicb.2015.01418

Granum, P. E. and Lund, T. (1997) '*Bacillus cereus* and its food poisoning toxins', *FEMS Microbiology Letters,* 157(2), pp. 223-228. DOI: 10.1111/j.1574-6968.1997.tb12776.x

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nat Meth,* 9(4), pp. 357-359. DOI: 10.1038/nmeth.1923

Leonard, S. R., Mammel, M. K., Lacher, D. W. and Elkins, C. A. (2015) 'Application of metagenomic sequencing to food safety: detection of Shiga toxin-producing

Escherichia coli on fresh bagged spinach', *Appl Environ Microbiol,* 81(23), pp. 8183-91. DOI: 10.1128/AEM.02601-15

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) 'The sequence alignment/map format and SAMtools', *Bioinformatics,* 25(16), pp. 2078-9. DOI: 10.1093/bioinformatics/btp352

Liu, Y., Lai, Q., Goker, M., Meier-Kolthoff, J. P., Wang, M., Sun, Y., Wang, L. and Shao, Z. (2015) 'Genomic insights into the taxonomic status of the *Bacillus cereus* group', *Sci Rep,* 5, pp. 14082. DOI: 10.1038/srep14082

Menzel, P., Ng, K. L. and Krogh, A. (2016) 'Fast and sensitive taxonomic classification for metagenomics with Kaiju', *Nat Commun,* 7, pp. 11257. DOI: 10.1038/ncomms11257

Miller, R. A., Kent, D. J., Boor, K. J., Martin, N. H. and Wiedmann, M. (2015a) 'Different management practices are associated with mesophilic and thermophilic spore levels in bulk tank raw milk', *J Dairy Sci,* 98(7), pp. 4338-51. DOI: 10.3168/jds.2015-9406

Miller, R. A., Kent, D. J., Watterson, M. J., Boor, K. J., Martin, N. H. and Wiedmann, M. (2015b) 'Spore populations among bulk tank raw milk and dairy powders are significantly different', *J Dairy Sci,* 98(12), pp. 8492-504. DOI: 10.3168/jds.2015-9943

Mtimet, N., Guegan, S., Durand, L., Mathot, A. G., Venaille, L., Leguerinel, I., Coroller, L. and Couvert, O. (2016) 'Effect of pH on *Thermoanaerobacterium thermosaccharolyticum* DSM 571 growth, spore heat resistance and recovery', *Food Microbiol,* 55, pp. 64-72. DOI: 10.1016/j.fm.2015.11.015

O-Thong, S., Prasertsan, P., Karakashev, D. and Angelidaki, I. (2008) 'Thermophilic fermentative hydrogen production by the newly isolated *Thermoanaerobacterium thermosaccharolyticum* PSU-2', *International Journal of Hydrogen Energy,* 33(4), pp. 1204-1214. DOI: 10.1016/j.ijhydene.2007.12.015

Peng, Y., Leung, H. C. M., Yiu, S. M. and Chin, F. Y. L. (2012) 'IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth', *Bioinformatics,* 28(11), pp. 1420-1428. DOI: 10.1093/bioinformatics/bts174

Pérez-Rodríguez, F., Valero, A., Carrasco, E., García, R. M. and Zurera, G. (2008) 'Understanding and modelling bacterial transfer to foods: a review', *Trends in Food Science & Technology,* 19(3), pp. 131-144. DOI: 10.1016/j.tifs.2007.08.003

Piro, V. C., Matschkowski, M. and Renard, B. Y. (2017) 'MetaMeta: integrating metagenome analysis tools to improve taxonomic profiling', *Microbiome,* 5(1), pp. 101. DOI: 10.1186/s40168-017-0318-y

Postollec, F., Mathot, A. G., Bernard, M., Divanac'h, M. L., Pavan, S. and Sohier, D. (2012) 'Tracking spore-forming bacteria in food: from natural biodiversity to selection by processes', *Int J Food Microbiol,* 158(1), pp. 1-8. DOI: 10.1016/j.ijfoodmicro.2012.03.004

Quigley, L., O'Sullivan, D. J., Daly, D., O'Sullivan, O., Burdikova, Z., Vana, R., Beresford, T. P., Ross, R. P., Fitzgerald, G. F., McSweeney, P. L., Giblin, L., Sheehan, J. J. and Cotter, P. D. (2016) 'Thermus and the pink discoloration defect in cheese', *mSystems,* 1(3). DOI: 10.1128/mSystems.00023-16

Rodrigues, M. X., Lima, S. F., Canniatti-Brazaca, S. G. and Bicalho, R. C. (2017) 'The microbiome of bulk tank milk: Characterization and associations with somatic cell count and bacterial count', *J Dairy Sci*. DOI: 10.3168/jds.2016-11540

Ronimus, R. S., Parker, L. E., Turner, N., Poudel, S., Ruckert, A. and Morgan, H. W. (2003) 'A RAPD-based comparison of thermophilic Bacilli from milk powders', *Int J Food Microbiol,* 85(1-2), pp. 45-61. DOI: 10.1016/s0168-1605(02)00480-4

Rudi, K., Moen, B., Dromtorp, S. M. and Holck, A. L. (2005) 'Use of ethidium monoazide and PCR in combination for quantification of viable and dead cells in complex samples', *Appl Environ Microbiol,* 71(2), pp. 1018-24. DOI: 10.1128/AEM.71.2.1018-1024.2005

Sadiq, F. A., Li, Y., Liu, T., Flint, S., Zhang, G. and He, G. (2016) 'A RAPD based study revealing a previously unreported wide range of mesophilic and thermophilic spore formers associated with milk powders in China', *Int J Food Microbiol,* 217, pp. 200-8. DOI: 10.1016/j.ijfoodmicro.2015.10.030

Scholz, M., Ward, D. V., Pasolli, E., Tolio, T., Zolfo, M., Asnicar, F., Truong, D. T., Tett, A., Morrow, A. L. and Segata, N. (2016) 'Strain-level microbial epidemiology and population genomics from shotgun metagenomics', *Nat Meth,* 13(5), pp. 435-438. DOI: 10.1038/nmeth.3802

Seale, R. B., Dhakal, R., Chauhan, K., Craven, H. M., Deeth, H. C., Pillidge, C. J., Powell, I. B. and Turner, M. S. (2012) 'Genotyping of present-day and historical *Geobacillus* species isolates from milk powders by high-resolution melt analysis of multiple variable-number tandem-repeat loci', *Appl Environ Microbiol,* 78(19), pp. 7090-7. DOI: 10.1128/AEM.01817-12

Setlow, P. (2003) 'Spore germination', *Current Opinion in Microbiology,* 6(6), pp. 550-556. DOI: 10.1016/j.mib.2003.10.001

Sharpton, T. J. (2014) 'An introduction to the analysis of shotgun metagenomic data', *Front Plant Sci,* 5, pp. 209. DOI: 10.3389/fpls.2014.00209

Silva, G. G., Green, K. T., Dutilh, B. E. and Edwards, R. A. (2016) 'SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data', *Bioinformatics,* 32(3), pp. 354-61. DOI: 10.1093/bioinformatics/btv584

Tallent, S. M., Kotewicz, K. M., Strain, E. A. and Bennett, R. W. (2012) 'Efficient isolation and identification of *Bacillus cereus* group', *J AOAC Int,* 95(2), pp. 446-51. DOI: 10.5740/jaoacint.11-251

Thoendel, M., Jeraldo, P., Greenwood-Quaintance, K. E., Yao, J., Chia, N., Hanssen, A. D., Abdel, M. P. and Patel, R. (2017) 'Impact of contaminating DNA in whole genome amplification kits used for metagenomic shotgun sequencing for infection diagnosis', *J Clin Microbiol*. DOI: 10.1128/JCM.02402-16

Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C. and Segata, N. (2015) 'MetaPhlAn2 for enhanced metagenomic taxonomic profiling', *Nat Meth,* 12(10), pp. 902-903. DOI: 10.1038/nmeth.3589

van Heel, A. J., de Jong, A., Montalbán-López, M., Kok, J. and Kuipers, O. P. (2013) 'BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides', *Nucleic Acids Research,* 41(Web Server issue), pp. W448-W453. DOI: 10.1093/nar/gkt391

Walsh, A. M., Crispie, F., Daari, K., O'Sullivan, O., Martin, J. C., Arthur, C. T., Claesson, M. J., Scott, K. P. and Cotter, P. D. (2017) 'Strain-level metagenomic analysis of the fermented dairy beverage nunu highlights potential food safety risks', *Applied and Environmental Microbiology*. DOI: 10.1128/aem.01144-17

Watterson, M. J., Kent, D. J., Boor, K. J., Wiedmann, M. and Martin, N. H. (2014) 'Evaluation of dairy powder products implicates thermophilic sporeformers as the primary organisms of interest', *J Dairy Sci,* 97(4), pp. 2487-97. DOI: 10.3168/jds.2013-7363

Weenk, G. H., van den Brink, J. A., Struijk, C. B. and Mossel, D. A. A. (1995) 'Modified methods for the enumeration of spores of mesophilic *Clostridium* species in dried foods', *International Journal of Food Microbiology,* 27(2), pp. 185-200. DOI: 10.1016/0168-1605(94)00164-2

Wehrle, E., Moravek, M., Dietrich, R., Burk, C., Didier, A. and Martlbauer, E. (2009) 'Comparison of multiplex PCR, enzyme immunoassay and cell culture methods for

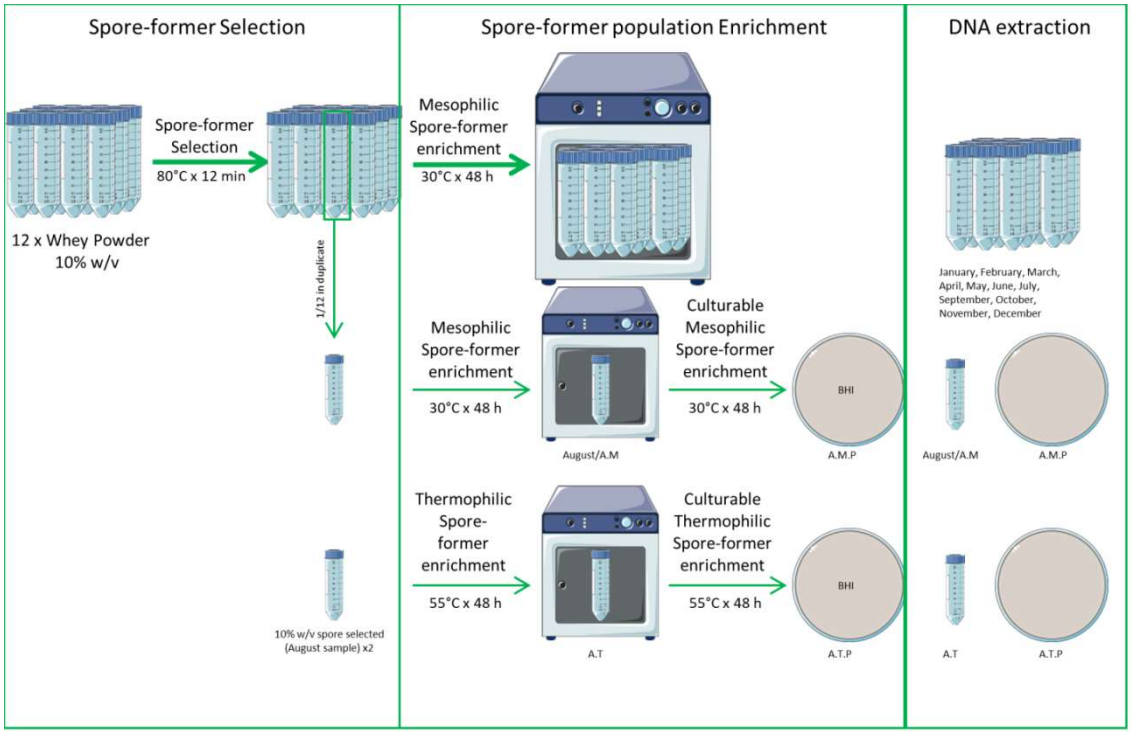
the detection of enterotoxinogenic *Bacillus cereus*', *J Microbiol Methods,* 78(3), pp. 265-70. DOI: 10.1016/j.mimet.2009.06.013

Wielinga, P. R., de Heer, L., de Groot, A., Hamidjaja, R. A., Bruggeman, G., Jordan, K. and van Rotterdam, B. J. (2011) 'Evaluation of DNA extraction methods for *Bacillus anthracis* spores spiked to food and feed matrices at biosafety level 3 conditions', *Int J Food Microbiol,* 150(2-3), pp. 122-7. DOI: 10.1016/j.ijfoodmicro.2011.07.023

Wood, D. E. and Salzberg, S. L. (2014) 'Kraken: ultrafast metagenomic sequence classification using exact alignments', *Genome Biology,* 15(3), pp. R46. DOI: 10.1186/gb-2014-15-3-r46

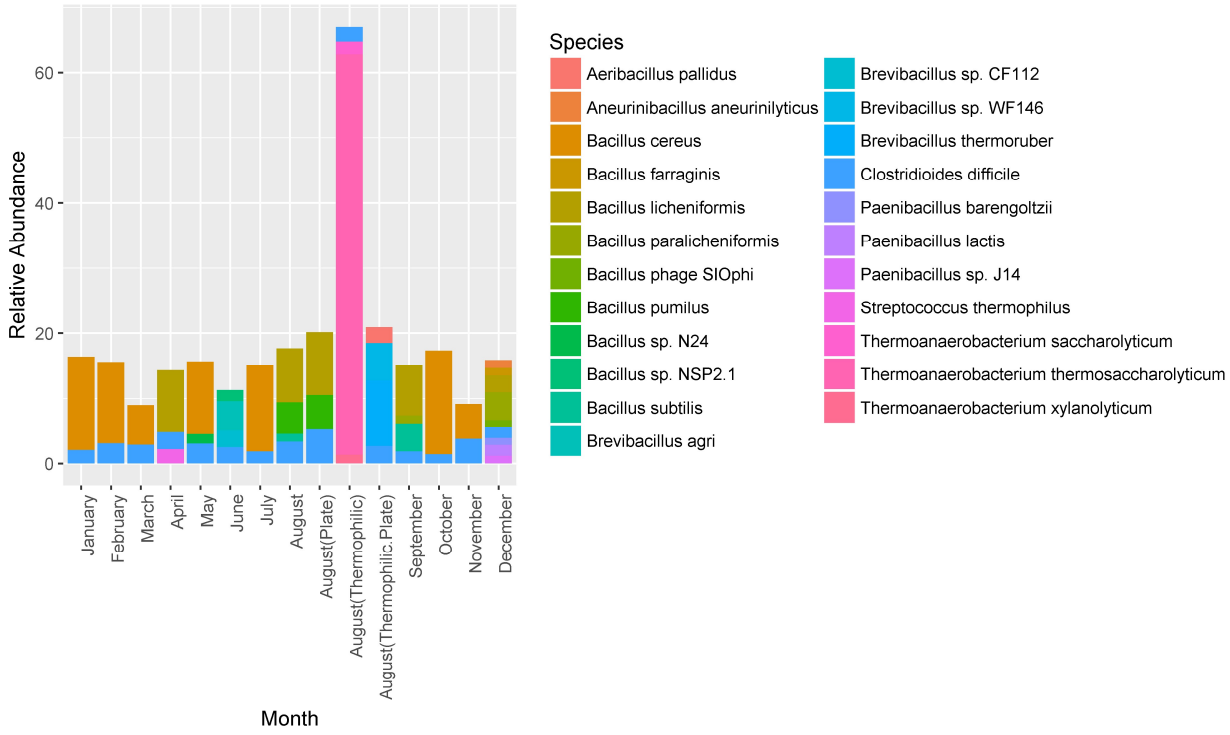
Yokouchi, H., Fukuoka, Y., Mukoyama, D., Calugay, R., Takeyama, H. and Matsunaga, T. (2006) 'Whole-metagenome amplification of a microbial community associated with scleractinian coral by multiple displacement amplification using phi29 polymerase', *Environ Microbiol,* 8(7), pp. 1155-63. DOI: 10.1111/j.1462-2920.2006.01005.x

3.9 Supplemental Figures



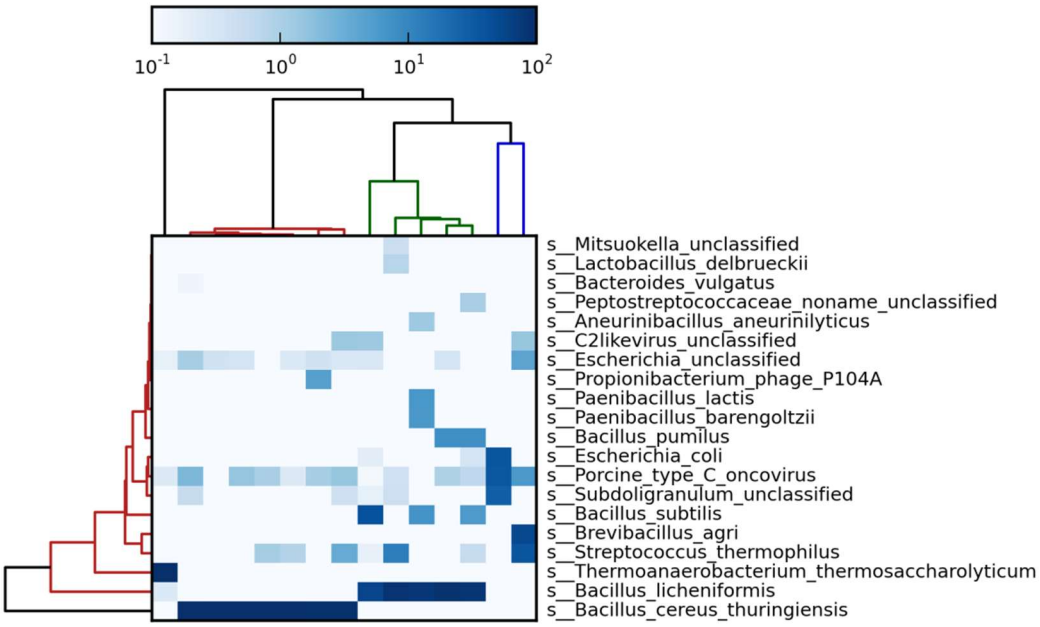
Supplemental Figure 3.1. Detailed schematic of spore selection and enrichment protocols followed by DNA extraction.

The schematic shows how each sample was treated for sporeformer selection and mesophilic sporeformer enrichment prior to DNA extraction. In addition one sample was also subject to thermophilic sporeformer enrichment (A.T), BHI culturable mesophilic sporeformer enrichment (A.M.P) and BHI culturable thermophilic sporeformer enrichment (A.T.P).



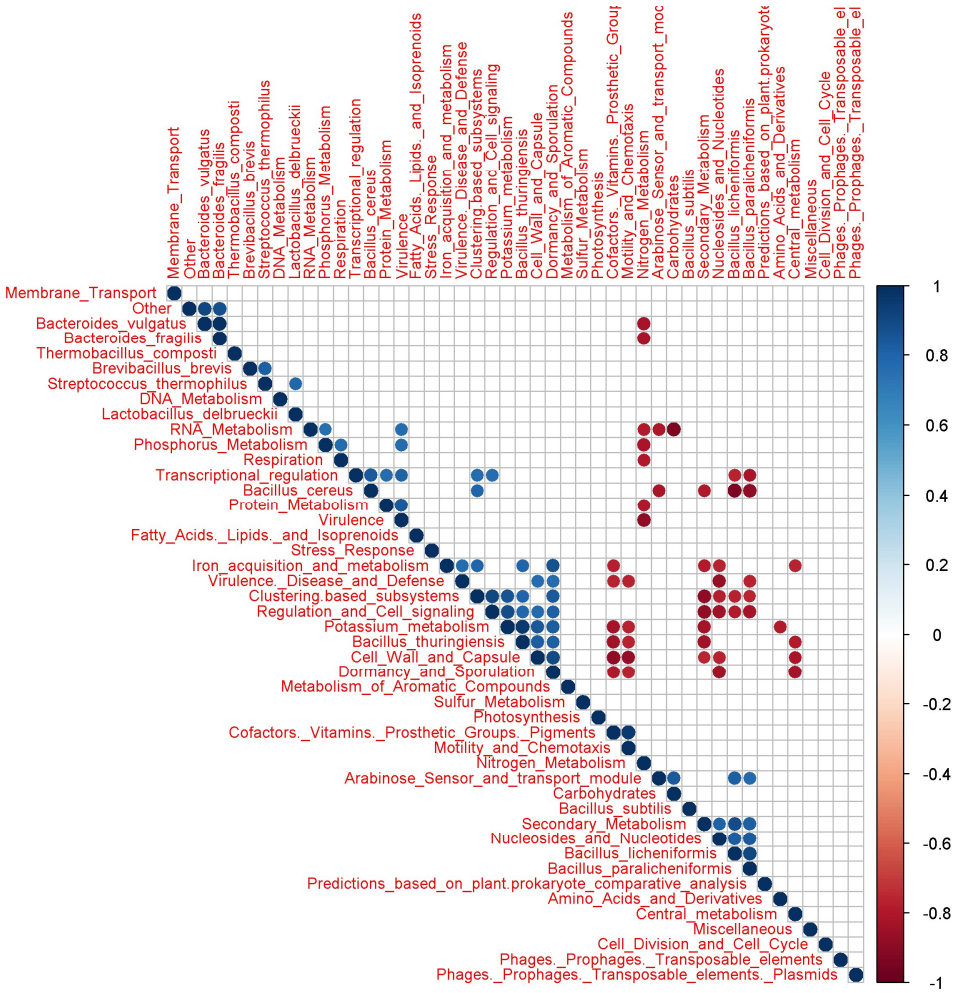
Supplemental Figure 3.2. Kaiju species taxonomy results.

Species were only included if they were present in > 1% relative abundance per sample.



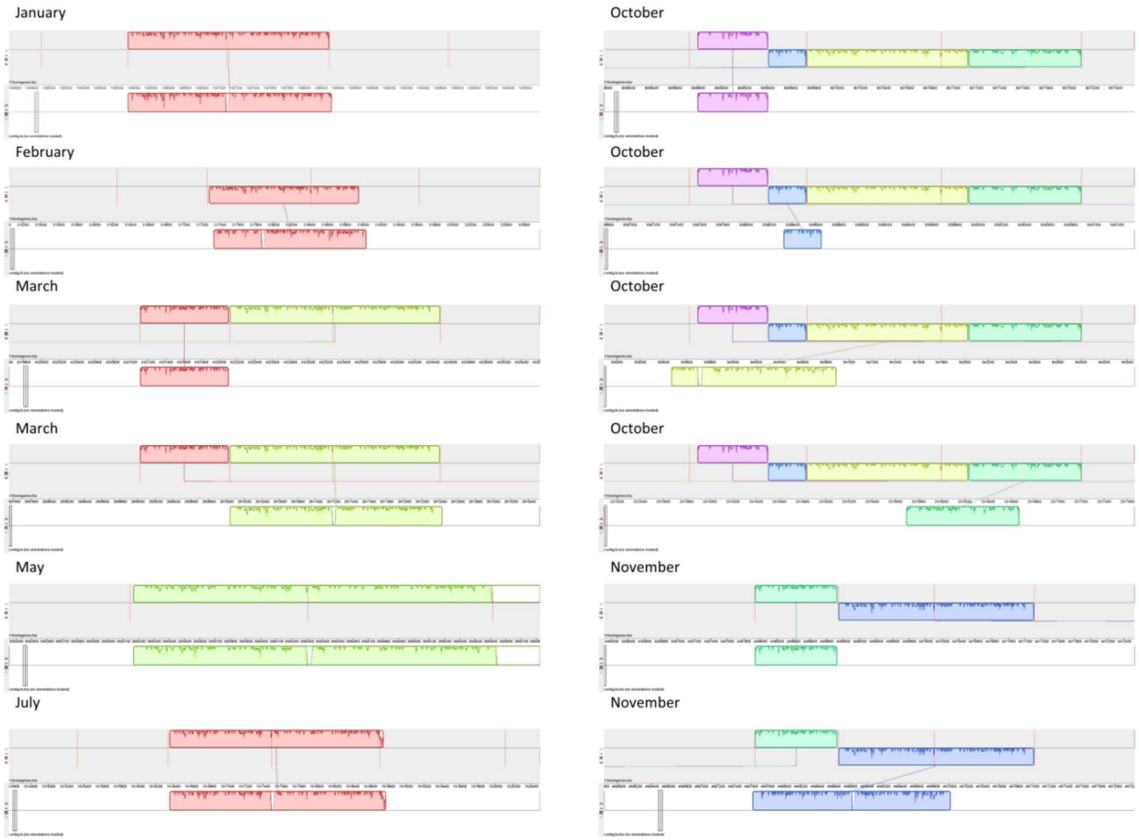
Supplemental Figure 3.3. Metaphlan species level classification.

From left to right samples are A.T, February, October, March, July, January, November, May, September, April, December, A.M.P, A.M, A.T.P, June.



Supplemental Figure 3.4. Spearman correlation with Benjamini-Hochberg correction for multiple comparisons between most abundant species identified and SUPER-FOCUS L1 functional groups.

Dots represent significant correlations $P \leq 0.05$. Blue being positive correlations and red being negative correlations.



Supplemental Figure 3.5. Mauve *B*. *cereus* toxin genes alignment.

3 toxin genes on 7 sets of contigs. Toxin genes on top row, CytK, Nhe L2 Nhe L1, contigs on bottom rows. Samples that align to more than one contig have multiple images. March and November have cytK gene on different contig to two NHE genes, so 2 alignments shown. October has 4 alignments shown as genes are at the edge of contigs.

Chapter 4. Tracking the dairy microbiota from farm bulk tank to skimmed milk powder

Authors: Aoife J. McHugh, Conor Feehily, Mark A. Fenelon, David Gleeson, Colin Hill, Paul D. Cotter

Contributions: Candidate performed DNA extraction, sequencing library preparation, bioinformatic, statistical analysis.

Candidate drafted and edited chapter.

DG organised sample collection.

All authors contributed, revised and approved manuscript.

4.1 Abstract

Microorganisms from the environment can enter the dairy supply chain at multiple stages including production, milk collection and processing, with potential implications for quality and safety. The ability to track these microorganisms can be greatly enhanced by the use of high throughput DNA sequencing (HTS). Here HTS, both 16S rRNA gene amplicon and shotgun metagenomic sequencing, were applied to investigate the microbiome of fresh, mid and late lactation milk collected from farm bulk tanks, collection tankers, milk silos, skimmed milk silos, a cream silo, and powder samples to investigate the microbial changes throughout a skim milk powder manufacturing process. 16S rRNA gene analysis established that the microbiota of raw milks from farm bulk tanks and in collection tankers were very diverse but that psychrotrophic genera associated with spoilage, *Pseudomonas* and *Acinetobacter*, were present in all samples. Upon storage within the whole milk silo at the processing facility, *Pseudomonas fluorescens* and *Acinetobacter baumannii* species became dominant. The skimmed milk powder generated during the mid lactation period had a microbial composition that was very different from that of raw milk, specifically, two thermophilic genera, *Thermus* and *Geobacillus*, were enriched. In contrast, the microbiota of skimmed milk powder generated from late lactation milk more closely resembled that of the raw milk, and was dominated by spoilage-associated psychrotrophic bacteria. This study demonstrates that the dairy microbiota can differ significantly across different sampling days. More specifically, HTS can be used to trace microbial species from raw milks through processing to final powdered products and has the potential to be used to develop a greater

understanding of the factors that dictate changes in the dairy microbiota throughout processing and, ultimately, lead to increased food safety and quality.

4.2 Introduction

Bovine milk is a nutritious natural food that can be processed into many different products, including dairy powders that can be used as a base for therapeutic, nutritional and/or infant formulas. Processing is required to provide a safe and stable shelf life and has a considerable impact on the microbial communities of the resultant products (Quigley *et al.*, 2013). With increased global demand for dairy products, including milk powders that are incorporated into infant milk formula, there is an even greater need to understand the associated microbiota in order to optimise food safety and quality. Such an understanding should incorporate an appreciation of the impact of both raw ingredients and processing environments on the final product (Quijada *et al.*, 2018; Yeluri Jonnala *et al.*, 2018; Fretin *et al.*, 2018; Cho *et al.*, 2018b; Wu *et al.*, 2018; Doyle *et al.*, 2017b). However, more information is needed to build a comprehensive view of the dairy microbiota and the factors that contribute to its composition. Traditional microbiological detection techniques focus on culturable bacteria. However, these approaches will not capture viable but non-culturable bacteria (Quigley *et al.*, 2013; Sadiq, Flint and He, 2018) or non-readily culturable bacteria (Zhao *et al.*, 2018). They can also be susceptible to false positives/negatives (Doyle, O'Toole and Cotter, 2018; Fricker, Reissbrodt and Ehling-Schulz, 2008), may not differentiate between closely related species, rely on a specific test for each individual target microbe, and are often time-consuming. More recently, high throughput DNA sequencing (HTS) has been considered as a means of determining the sources of spoilage and pathogenic bacteria in raw milk. These methods have been used to study the influence of environmental factors,

including animal housing, cleaning and milking practices (Doyle *et al.*, 2017b), the weather conditions (Li *et al.*, 2018), seasonal influences, and on farm storage conditions (Doyle *et al.*, 2017a), as well as large scale storage and seasonal transportation (Kable *et al.*, 2016), on the raw milk microbiota. Seasonality can be particularly important in pasture-based systems when milk quality is impacted by where the herd in located (indoor vs. outdoor) amongst other factors. In addition, associations between the raw milk microbiota and somatic cell counts (a hygiene and herd health indicator) in bulk tank samples (Rodrigues *et al.*, 2017; Li *et al.*, 2018) as well as the impacts of pasteurisation on the milk microbiota (Quigley *et al.*, 2013) have been studied using these HTS methods. These studies are limited by the use of 16S rRNA gene amplicon sequencing, which can provide taxonomic resolution to the genus level only. Therefore, there is an insufficient understanding of the functional potential of the microbial communities and indeed characterisation of the non-bacterial microbiome along the dairy chain. Recently shotgun metagenomic sequencing, which overcomes these issues, has been used to study dairy products (Walsh *et al.*, 2017; Wolfe *et al.*, 2014; Quigley *et al.*, 2016). Here, 16S rRNA gene amplicon and shotgun metagenomic analysis are used together to facilitate an in-depth study of the dairy microbiome from the farm, through transportation and processing to a skimmed milk powder.

4.3 Materials and Methods

4.3.1 Sample collection

Raw milks, pasteurised milks and powdered dairy products were sampled from within a commercial milk processing pipeline (Figure 4.1). Raw milk bulk tanks were sampled on one day during the early-to-mid lactation period (May 2016), hereafter referred to as mid lactation. These milks were combined and further processed to a skimmed milk powder. During this process, samples were also collected from collection tankers, the processor's whole milk silo, the skimmed milk silo and the resultant skimmed milk powders. Samples from raw milk bulk tanks (October 2016) and the processing pipeline (December 2016), which in this instance also included sampling of the cream silo, were also collected later in the year to represent the late lactation period. In order to complete the full process a minimum capacity of milk was required, considering lower herd production in late lactation a greater number of farm bulk tanks were required. Therefore, during the late lactation period, the original farms were resampled separately from the process (collection tanker to skimmed milk powder), which contained different farm bulk tank inputs (not sampled). Samples were collected by personnel at the processing facility using standard collection procedures and liquid samples were transported on ice to the laboratory, where DNA extraction was performed immediately. Skimmed milk powders were transported at room temperature and stored for up to 1 month prior to DNA extraction.

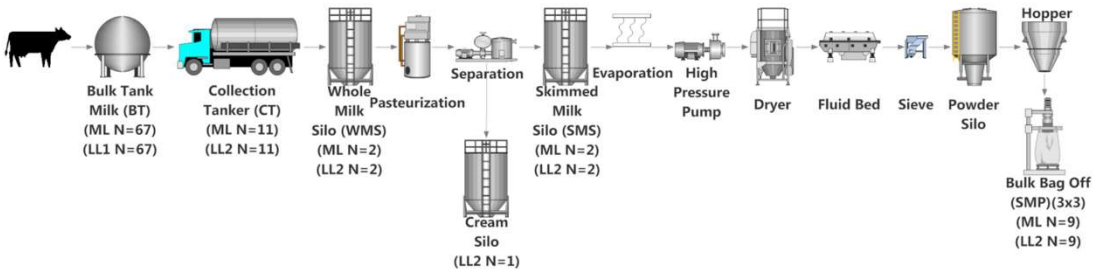


Figure 4.1. Sampling schematic.

Sampling included mid lactation samples (ML = May 2016) where a whole processing run was sampled from farm bulk tank milk to skimmed milk powder. This included 67 farm bulk tanks (BT), the 11 collection tankers (CT) used to collect this milk and the whole milk silo (WMS) into which this milk was pooled (2 samples were obtained). Milk in the WMS was subject to pasteurisation and separation and stored in a skimmed milk silo (SMS) from which 2 samples were obtained. Milk in the SMS was subject to heating and drying to make a skimmed milk powder (SMP), from which 9 samples were obtained, 3 samples from each of 3 bags. The same 67 farms were resampled during late lactation (October 2016; Late Lactation 1 = LL1). On a separate day during this late lactation period (December 2016; Late Lactation 2 = LL2), samples from tankers and the processing run, including an additional cream sample, were collected for analysis.

4.3.2 DNA extraction

DNA was extracted from fresh liquid dairy samples. Powders were stored at room temperature and, upon reconstitution at 10% w/v in ¼ strength Ringers solution, DNA was extracted from 30 ml milk samples and 50 ml reconstituted skimmed milk powder samples using the MoBio PowerFood DNA Isolation kit as directed within the manufacturer's instructions with some minor adjustments. Samples were centrifuged at 5000 g for 20 min at 4°C. Fat was removed (from raw milk samples) and supernatant discarded. From here pellets were washed, and subject to lysozyme treatment as previously reported (McHugh *et al.*, 2018). Twenty eight microliters of proteinase K was added and incubated at 55°C for 15 min. Samples were centrifuged at 13,000 g for 1 min and supernatant discarded. Pellets were resuspended in 450 µl PF1 from the PowerFood kit and, from this point, the kit protocol was followed, including the recommended alternative lysis step for difficult-to-lyse cells. DNA was eluted in 50 µl elution buffer.

4.3.3 16S rRNA Amplicon sequencing

Template DNA was quantified and checked for quality using the Qubit dsDNA High Sensitivity Assay kit as well as running 2 µl on a 1% agarose gel. DNA was normalised to 5 ng μl^{-1}. The V3-V4 variable region of the 16S rRNA genes was amplified in triplicate from each sample as previously described (Doyle *et al.*, 2017b) with a few changes; 35 PCR cycles were used instead of 32, and KAPA2G Robust(Kapa Biosystems Ltd.) was used instead of Kapa HiFi Hotstart. Two microliters of each PCR reaction was ran on 1% agarose gel to check for quality before pooling triplicate PCR reactions, cleaning with 0.8 x volume Ampure XP

beads. Fifty microliters of cleaned up sample was stored at -20°C. Five microliters was aliquoted and subject to index PCR and clean up according to Illumina 16S Metagenomic Sequencing Library Preparation guidelines as previously described (Doyle *et al.*, 2017b). DNA concentrations were quantified using Qubit dsDNA High Sensitivity Assay kit before diluting to 20 nM, pooling and performing a final 1:1 Ampure XP clean up. The samples were pooled into 4 pools. Samples from each processing step were contained in each pool, with mid lactation skimmed milk powder samples included in each pool as a control. The samples were sequenced on Illumina MiSeq sequencing platform in the Teagasc sequencing facility using 2 x 250 V2 kit, according to Illumina sequencing protocols.

4.3.4 Whole metagenome shotgun sequencing

A subset of samples were selected for whole metagenome shotgun sequencing. These were 4 WMS (2 x mid lactation, 2 x late lactation), 4 SMS (2 x mid lactation, 2 x late lactation), 6 SMP (3 x mid lactation, 3 x late lactation (1 from each bag)), as well as 2 pooled collection tanker (CT_P) samples, in which equal volumes of DNA from each of the 11 mid lactation samples were pooled into one sample and equal volumes of each of the late lactation tankers were pooled into a second sample. Samples were prepared according to the Nextera XT DNA library preparation guide from Illumina. Samples were sequenced on Illumina MiSeq sequencing platform at Teagasc sequencing facility with a 2 x 250 V2 kit with standard Illumina sequencing protocols.

4.3.5 Bioinformatic and statistical analysis

16S rRNA gene amplicon sequences were processed as previously described (Doyle *et al.*, 2017b). Briefly, forward and reverse reads were joined using usearch FLASH (fast length adjustment of short reads to improve genome assemblies) (Magoc and Salzberg, 2011). Paired end reads were further processed by quality filtering based on quality score of 25 and removing mismatched barcodes and sequences below length thresholds by QIIME (Caporaso *et al.*, 2010b). USEARCH v7 (64-bit) (Edgar, 2010) was utilised for removing noisy data, chimera detections and clustering into operational taxonomic units (OTUs) at 97% identity. OTUs were aligned using PyNAST (python nearest alignment space termination; a flexible tool for aligning sequences to a template alignment; (Caporaso *et al.*, 2010a)) and taxonomy was assigned using BLAST (Altschul *et al.*, 1990) against the SILVA SSURef database release123 (Quast *et al.*, 2013). Qiime data was further analysed using Phyloseq in R (McMurdie and Holmes, 2013) richness was estimated to get alpha diversty, and distances obtained for beta diversity using phyloseq before visualising using ggplot2 (Wickham, 2009). Taxonomy was also visualised using ggplot2. Pairwise wilcoxon rank sums test using Benjamini hochberg *p*-value correction/fdr correction analysis was used to compared samples groups from mid and late lactation. Adonis from the vegan R package was used to determine differences in beta diversity. Shotgun metagenomic data was processed as previously described (McHugh *et al.*, 2018). Briefly, raw metagenomic reads were checked for the presence of bovine reads which were removed, filtered on the presence of quality and quantity, and trimmed to 170 bp with a combination of Picard tools and SAMtools (Li *et al.*, 2009). Kraken

with a filter threshold of 0.2 (Wood and Salzberg, 2014) and SUPER-FOCUS (Silva *et al.*, 2016) were used to determine microbial composition to species level and biological functions, respectively.

4.4 Results

4.4.1 Bulk tank milks contain a diverse microbiota that differs in samples collected from the mid and late lactation periods

DNA was extracted and 16S rRNA gene amplicons sequenced for a total of 67 raw bulk tank milk samples (BT) collected on one day in the early-to-mid lactation period (May 2016), hereafter referred to as mid lactation. This process was repeated in the late lactation period (October 2016) (Figure 4.1). The alpha diversity of the raw milk microbiota in bulk tanks on farms was relatively high (Figure 4.2A), compared to subsequent processing stages. However, the microbial alpha diversity in mid lactation bulk tank samples was significantly lower ($p < 0.001$) than the corresponding late lactation samples (Figure 4.2A). Beta diversity showed that samples from bulk tank milks were dissimilar but broadly clustered together (Figure 4.2B).

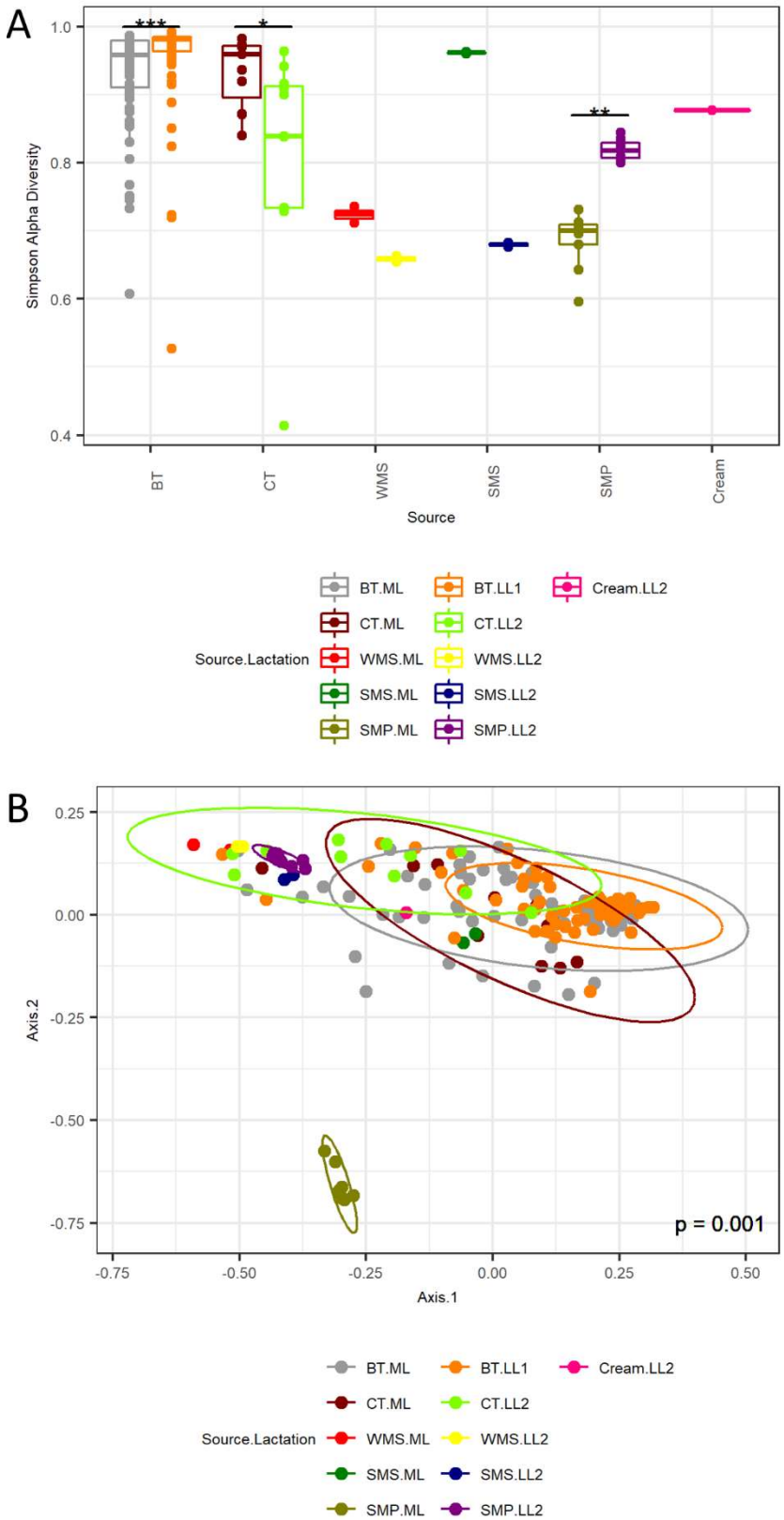


Figure 4.2. Microbial diversity indexes of sampling sites.

[Figure 4.2 continued]

A. Simpson alpha diversity analysis of 16S rRNA gene amplicon sequence data for mid (ML) and late lactation samples (LL1/LL2) from bulk tanks (BT), collection tankers (CT), whole milk silos (WMS), skimmed milk silos (SMS), skimmed milk powder (SMP) and cream (LL2 only) samples. Significant differences are highlighted (*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

B. Bray Curtis multi-dimensional scaling analysis of 16S rRNA gene amplicon sequence data for mid (ML) and late lactation samples (LL1/LL2) from bulk tanks (BT), collection tankers (CT), whole milk silos (WMS), skimmed milk silos (SMS), skimmed milk powder (SMP) and cream (late lactation only) samples.

Farm bulk tanks were composed of a high number of genera that were present at low relative abundance of less than 5%. In one sample these low abundance genera accounted for 74.9% of the total population and an average of 46.4% was seen across all bulk tank samples (Figure 4.3, Supplemental Figure 4.1, Supplemental Table 4.1). There were 42 genera present at high relative abundance (> 5% relative abundance in at least one sample) (Supplemental Table 4.1) including traditionally milk-associated taxa such as *Pseudomonas* (mean 6.6%), *Acinetobacter* (mean 5.2%), *Lactococcus* (mean 4.7%), *Corynebacterium* (mean 4.2%) and *Streptococcus* (mean 2.5%) (Figure 4.3, Supplemental Table 4.1). In general, the microbial diversity of the bulk tank milks was such that taxonomic composition differed across farms and the two sampling days. However, it was apparent that the bulk tank milk profiles from some farms remained relatively more stable across the two sampling points, e.g., Farm 23 had a high relative abundance of *Leuconostoc* and *Acinetobacter*, in both the mid and late lactation samples (Figure 4.3).

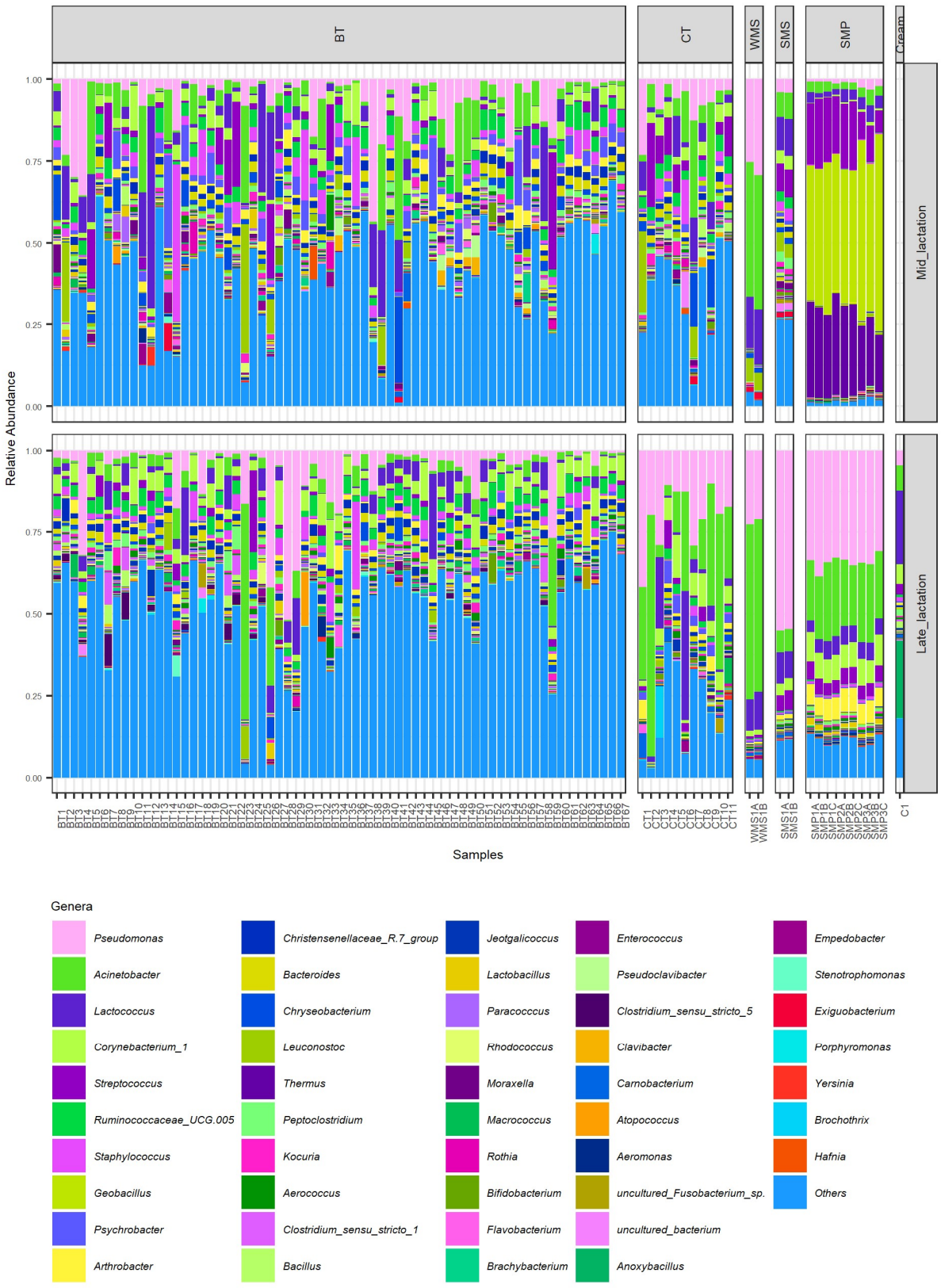


Figure 4.3. Relative abundance of genera present in > 5% relative abundance in at least one sample.

[Figure 4.3 continued]

16S rRNA gene amplicon sequence data for mid and late lactation samples from bulk tanks (BT), collection tankers (CT), whole milk silos (WMS), skimmed milk silos (SMS), skimmed milk powders (SMP) and cream (late lactation only) samples. Genera shown are present in > 5% relative abundance in at least one sample.

Overall, 17 high relative abundance genera, which were genera present at greater than 5% relative abundance in at least one sample, differed significantly in their relative abundance between paired mid and late lactation bulk tank samples ($p < 0.05$). Nine were higher in relative abundance in mid lactation bulk tanks, whilst 8 were more abundant in late lactation bulk tanks (Figure 4.4). Amongst these, genera grouped as *Clostridium sensu stricto* 1 and 5 were both present at significantly higher relative abundance in the late lactation bulk tank milks. Low abundance genera, grouped together as others, were also detected at significantly greater proportions in late lactation bulk tanks (Figure 4.4). One mid lactation sample had a particularly high proportion of *Yersinia* (5.6%), leading to mid lactation bulk tanks having, on average, a significantly higher relative abundance of this genus. There were also a larger number of genera (77) that were present at a lower relative abundance, but greater than 1% abundance in at least one sample, that significantly differed in abundance between mid and late lactation bulk tank samples (Supplemental Figure 4.2). It was again notable that taxa corresponding to the genus *Clostridium*, in this instance *sensu stricto* 15 and 18, were detected at significantly higher relative abundance in late lactation samples.

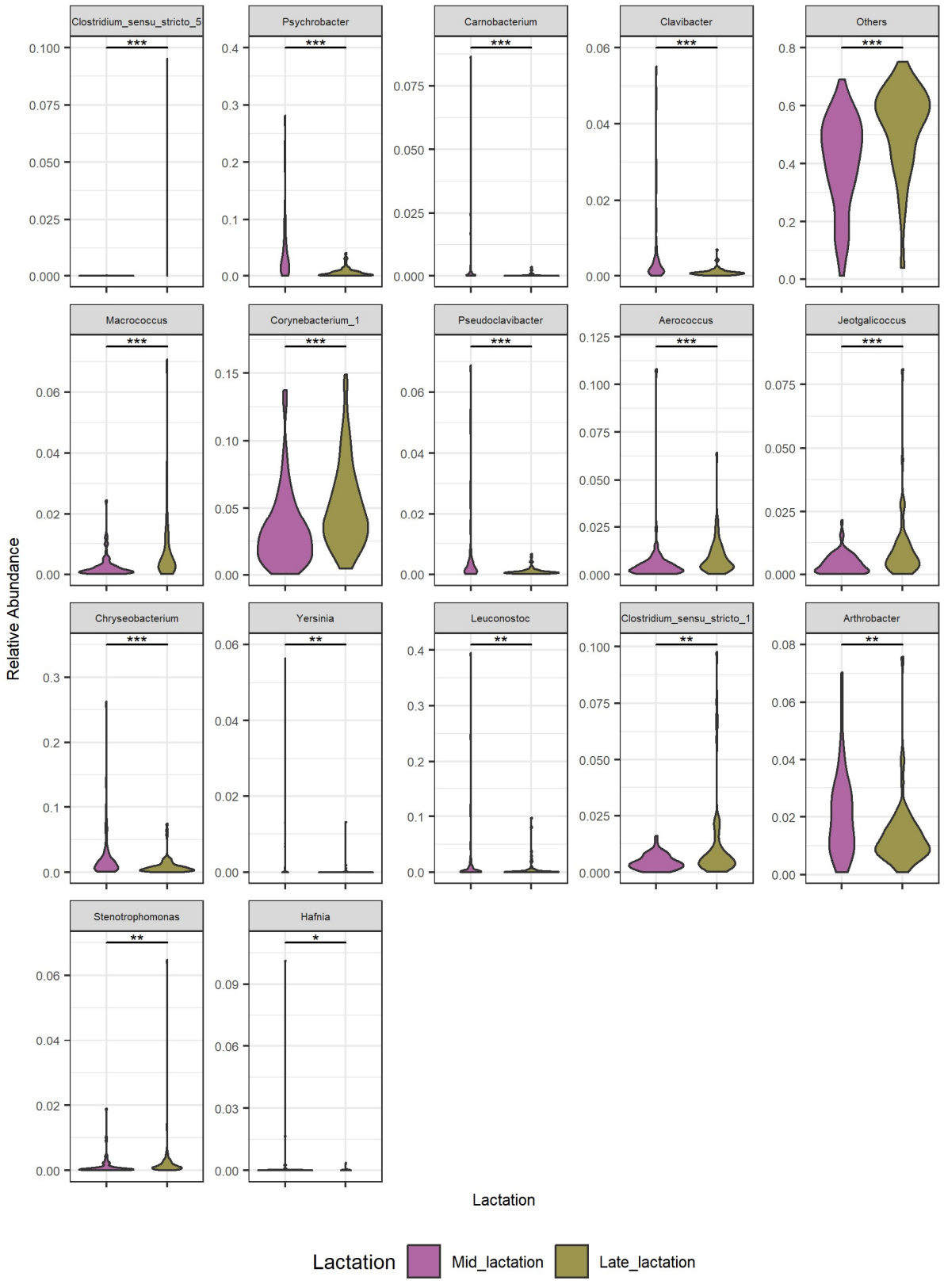


Figure 4.4. Significantly differential relative abundance of taxa constituting at least 5% relative abundance of at least one sample between mid and late lactation bulk tanks.

[Figure 4.4 continued]

Violin plots of genera that differ significantly in relative abundance between mid and late lactation bulk tanks. This highlighted genera that are present in > 5% relative abundance in at least one sample. Pairwise wilcoxon rank sums test using Benjamini hochberg *p*-value correction/fdr correction analysis was performed and samples ordered by lowest *p*-value to highest (*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

4.4.2 Collection tanker milks retain a relatively high microbial diversity but with some taxonomic convergence

Milk samples obtained from collection tankers (CT) were, with the exception of bulk tank milks, among containing the highest microbial alpha diversity (Figure 4.2A). Milk from mid lactation collection tankers had a significantly higher microbial alpha diversity ($p < 0.05$) compared to the corresponding late lactation samples (Figure 4.2A). From a beta diversity perspective, mid lactation CT samples cluster closely to the bulk tanks from which they were filled (Figure 4.2B). Late lactation CT samples cluster further away from bulk tank samples, reflecting their collection on different days (as, in the late lactation period, the 67 farms from which bulk tank milk was collected did not yield a sufficiently large pool of milk to proceed with the powder manufacturing process) (Figure 4.2B). Although the microbial composition of the tanker samples was diverse, and no two tankers had the same composition, a pattern of enrichment was apparent with respect to the taxa present at highest relative abundance (Supplemental Table 4.1, Figure 4.3), with *Acinetobacter*, *Pseudomonas*, *Lactococcus* and *Corynebacterium* on average accounting for between 9% and 3% relative abundance of mid lactation tanker milk microbiota composition (Figure 4.3). Analysis of the corresponding samples collected during late lactation showed that although the tanker's milk microbiota composition was diverse, *Pseudomonas* (mean 20.3%) and *Acinetobacter* (mean 25.4%) became the most dominant genera in each sample (Figure 4.3, Supplemental Table 4.1). Mid lactation tanker milk samples had a greater proportion of low (< 5%) and very low (< 1%) abundance genera, compared to late lactation tanker samples (Figure 4.3,

Supplemental Figure 4.1). *Pseudomonas*, *Bacillus* and *Ruminococcaceae* UCG.005 differed significantly between mid and late lactation tanker samples, with *Pseudomonas*, or more specifically *Pseudomonas fluorescens* as determined through shotgun metagenomic sequencing, at higher relative abundance in late lactation samples, and *Bacillus* and *Ruminococcaceae* UCG.005 at higher relative abundance in mid lactation samples (Figure 4.5). With respect to *Bacillus*, the pattern was driven by higher relative abundance in mid lactation tankers 4, 8 and 11 (Figure 4.3, Supplemental Figure 4.1). Shotgun metagenomic data established that, at the species level, this taxon corresponded to *Bacillus coagulans* (Figure 4.6). When SUPER-FOCUS was used to assign functional classification to shotgun metagenomic reads (Supplemental Figure 4.3), a higher relative abundance of virulence functions were noted in raw milks in pooled tanker samples than heat processed samples of skimmed milk silos or skimmed milk powders (Figure 4.7). A high proportion of the shotgun metagenomic reads sequenced were not of microbial origin and assigned to *Bos taurus* (Supplemental Figure 4.4) and so the resulting low number of microbial associated reads did not allow strain level classification or in depth functional classification.

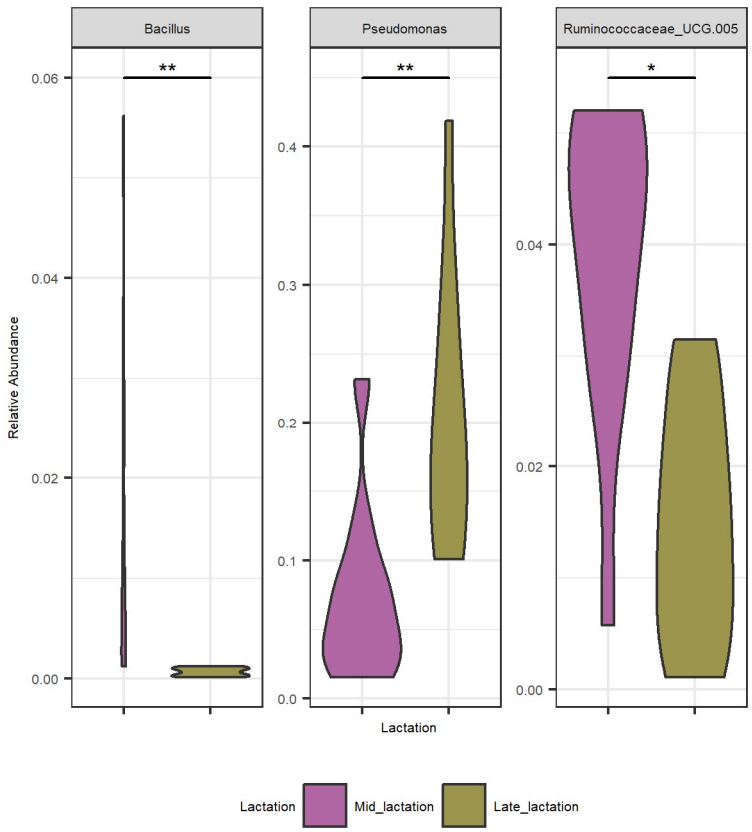


Figure 4.5. Significantly differential relative abundance of taxa constituting at least 5% relative abundance in at least one sample between mid and late lactation collection tankers.

Violin plots of genera that differ significantly in relative abundance between mid and late lactation collection tankers. This highlights genera that are present in > 5% relative abundance in at least one sample. Pairwise wilcoxon rank sums test using Benjamini hochberg *p*-value correction/fdr correction analysis was performed and samples ordered by lowest *p*-value to highest (*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

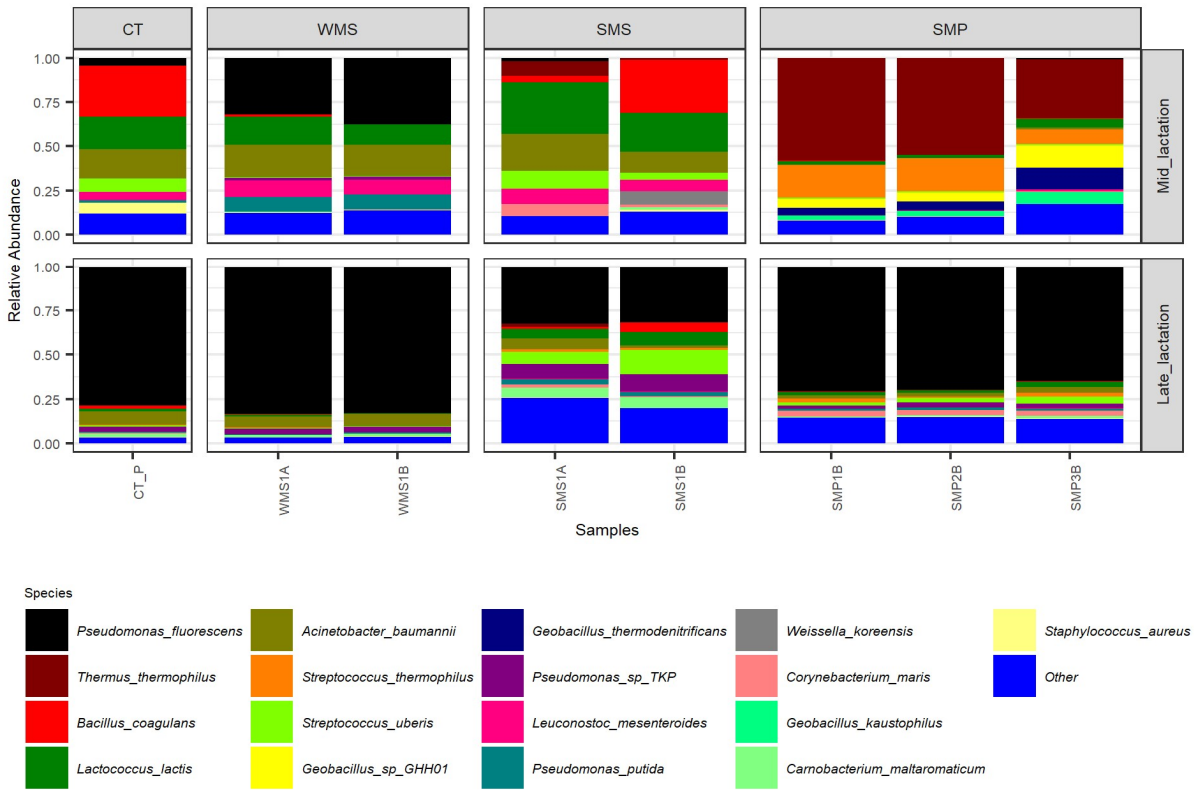


Figure 4.6. Relative abundance of species present in > 5% relative abundance in at least one sample.

Shotgun species level taxonomic classification from Kraken analysis using filter threshold 0.2 on a subset of samples from the mid and late lactation processing pipelines, including a pooled representative collection tanker sample (CT_P), whole milk silos (WMS), skimmed milk silos (SMS) and a subset of skimmed milk powder samples (SMP) from each lactation stage.

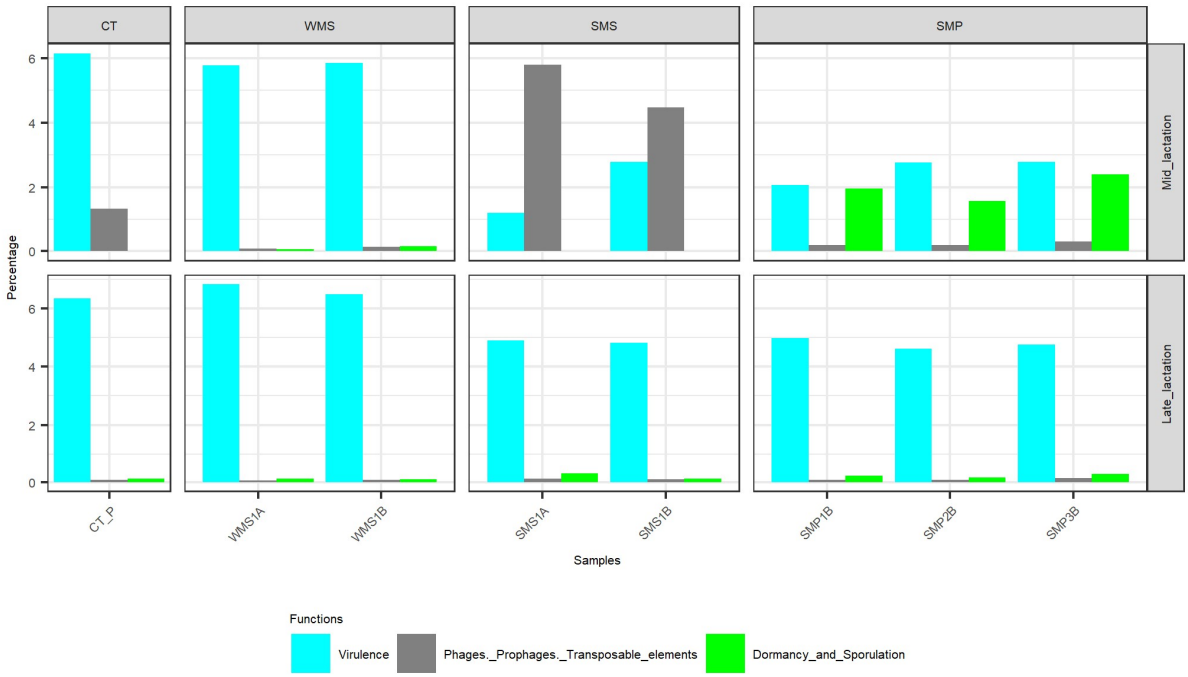


Figure 4.7. Percent of a subset of functional gene groups.

Percent of a subset of functional classifications from SUPER-FOCUS L1 analysis attributed to prophage, phage and transposable elements, dormancy and sporulation as well as virulence, in mid and late lactation collection tanker pooled sample (CT_P), whole milk silo (WMS), skimmed milk silo (SMS) and skimmed milk powder (SMP).

4.4.3 Whole milk silo storage results in the increased dominance of psychrotrophic spoilage-associated bacteria

A decrease in microbial alpha diversity was observed in whole milk silo (WMS) samples, regardless of the period of collection compared to the corresponding samples from collection tankers (Figure 4.2A). Beta diversity analysis displayed tight clustering of whole milk silo samples, regardless of lactation stage. Whole milk silo samples cluster together, away from mid lactation collection tankers and at the edge of the ellipse for late lactation collection tankers (Figure 4.2B). A high relative abundance of spoilage-associated psychrotrophic bacteria was observed in both mid and late lactation samples (Figure 4.3), with *Pseudomonas*, *Acinetobacter*, as well as the non-psychrotrophic *Lactococcus* genus, dominating. Taxonomic composition from mid and late lactation samples was similar (Figure 4.3), *Leuconstoc* was the only genus present at a higher abundance in mid lactation (mean; 6.3%), compared to late lactation samples (mean; 0.1%) (Supplemental Table 4.1). In contrast, low abundance genera, denoted as 'Others', were present at a higher relative abundance in late lactation (mean; 5.4%), than mid lactation (mean; 3.1%) samples (Supplemental Table 4.1). However, all WMS samples had a smaller proportion of low abundance genera compared to collection tankers and bulk tank milks (Figure 4.3). This reflected the lower alpha diversity of these samples relative to those collected earlier in the milk processing chain (Figure 4.2A). Shotgun metagenomic analysis revealed species level classification of the most abundant genera across all WMS samples as *Pseudomonas fluorescens*, *Acinetobacter baumannii* and *Lactococcus lactis* (Figure 4.6). Due to the high

proportion of shotgun metagenomic reads assigned to *Bos taurus* (Supplemental Figure 4.4), the resulting low number of bacterial associated reads did not allow strain level classification. When SUPER-FOCUS was used to assign functional classification to the shotgun metagenomic reads (Supplemental Figure 4.3), a higher relative abundance of virulence functions were noted in raw milks in the WMS and pooled tankers than all other samples (Figure 4.7).

4.4.4 The skimmed milk silo microbial composition differs across samples collected in different seasons

The alpha diversity of the skimmed milk silo (SMS) microbiota was greater than that of the preceding whole milk silo samples (Figure 4.2A). Beta diversity analysis showed that the microbiota of mid lactation SMS samples cluster separately from WMS samples. In contrast, late lactation samples clustered closely with their corresponding WMS samples (Figure 4.2B). From a taxonomic perspective, the dominant psychrotrophic bacteria detected at the preceding WMS stage was reduced in relative abundance in the mid lactation SMS samples following pasteurisation and separation of that milk (Figure 4.3). This decrease was not as evident in late lactation samples, with *Pseudomonas*, *Lactococcus* and *Acinetobacter* remaining dominant (Figure 4.3). SMS samples from both time points contained *Lactococcus*, *Streptococcus* and *Corynebacterium* (Figure 4.3). Species level taxonomic analysis revealed that these primarily corresponded to *Lactococcus lactis*, *Cornyebacterium maris*, *Streptococcus thermophilus* and *Streptococcus liberis* at the species level (Figure 4.6). Mid lactation SMS samples had a greater proportion of low and very low abundance bacteria compared to late

lactation SMS (Figure 4.3, Supplemental Table 4.1, Supplemental Figure 4.1). When SUPER-FOCUS was used for functional classification of shotgun metagenomic reads (Supplemental Figure 4.3), a higher relative abundance of genes associated with phage, prophage and transposable elements were noted in mid lactation SMS samples compared to all other samples (Figure 4.7).

4.4.5 The cream silo microbiota differs from that of other dairy processing samples

The late lactation processing pipeline provided the only cream sample available for analysis within this study. This sample had a higher alpha diversity than the whole milk silo sample from which it was produced, and the skimmed milk silo contents, from which it was separated (Figure 4.2A). The microbial communities in the sample did not cluster with either the late lactation skimmed silo or the whole milk silo samples (Figure 4.2B). This sample was dominated by *Anoxybacillus* (Figure 4.3), a genus present at < 1% relative abundance at previous stages of the processing pipeline, as well as *Lactococcus*, *Corynebacterium*, *Pseudomonas* and *Acinetobacter* amongst other genera (Supplemental Table 4.1, Figure 4.3). Low abundance genera accounted for 18.0% of the cream silo microbiota (Supplemental Table 4.1) and 6.4% of the microbiota corresponded to very low abundance genera.

4.4.6 The dairy powder microbiota can vary; reflecting either the original raw milk microbiota or microbes selected for during processing

The microbiota of skimmed milk powder samples (SMP) differed in a manner that reflected the processing days. The microbiota of mid lactation skimmed milk powder samples had a lower alpha diversity compared to the SMS milk from which

it was produced (Figure 4.2A). Furthermore, the microbial communities in the mid lactation skimmed milk powders clustered separately from both the late lactation skimmed milk powder samples and the milks from which they were derived (Figure 4.2B). More specifically, the mid lactation powders showed a shift in microbial taxonomic dominance with thermophilic bacteria, such as *Thermus*, *Geobacillus* and *Streptococcus*, being more dominant in these samples (Figure 4.3). Shotgun metagenomic sequencing assigned these as *Thermus thermophilus*, *Geobacillus GHH01*, *Geobacillus thermodenitrificans* and *Streptococcus thermophilus*. SUPER-FOCUS functional classification highlighted a corresponding increase in the relative abundance of sporulation and dormancy-associated genes in mid lactation skimmed milk powder samples (Figure 4.7), reflecting the proportions of spore-forming bacteria present. In these samples, low relative abundance genera (< 5%) accounted for 0.9% - 2.9% of reads and very low abundance genera (< 1%) accounted for 0.45% - 1.3% of reads. The microbiota of late lactation skimmed milk powders differed considerably. An increase in diversity was observed in late lactation skimmed milk powders compared to the SMS milk from which they were produced (Figure 4.2A) and these samples clustered closely to the skimmed milk samples from which they were derived (Figure 4.2B). Taxonomic analysis revealed the dominance of psychrotrophic genera and, more specifically, of the same genera and species (*P. fluorescens* and *A. baumannii*) that had dominated previous WMS and SMS samples (Figure 4.3, Figure 4.5). Late lactation powder had significantly higher ($p < 0.01$) microbial alpha diversity compared to mid lactation powder (Figure 4.2A).

4.4.7 Sample source and lactation stage significantly influence differences between the microbiota of the dairy samples

Overall, Adonis analysis from the R vegan package of Bray Curtis beta diversity analysis showed significant differences between samples based on lactation stage and source of sample ($p \leq 0.001$), Twenty percent of the variation in distance between samples was attributed to sample source (BT/CT/WMS/SMS/SMP/Cream), 9.5% of the variation in distance between samples was to lactation stage(ML, LL1, LL2), with 4.7% due to both source and lactation stage.

4.5 Discussion

This study set out to use molecular methods to provide an important description of the microbiota of a food processing pipeline, by tracking the microbiota of raw milks on farms to a final skimmed milk powder. Through HTS, it was demonstrated that different production days and microbial selection by process parameters can impact the microbiota during the process (Figure 4.8). Through this approach the study expands upon previous investigations (Kable *et al.*, 2016; Alessandria *et al.*, 2016; Parente *et al.*, 2016; Stellato *et al.*, 2017) to give an even greater understanding of the changes in the dairy microbiota from milk to skimmed milk powder. This investigation confirms the diversity of the raw milk microbiota, however it highlights that bulk tank samples from mid and late lactation broadly cluster together.

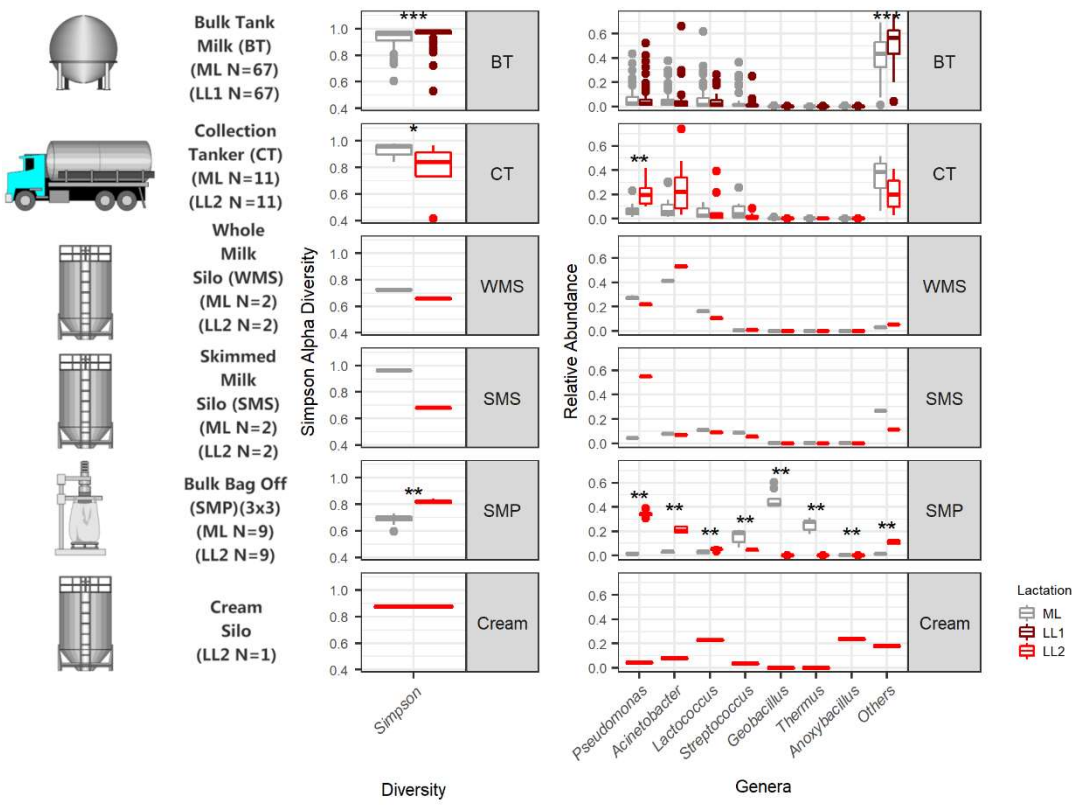


Figure 4.8. Summary of results.

Summary of differences due to sample location and lactation stage. Brief sampling schematic, with N numbers, Simpson alpha diversity as well as genera with an average of at least 10% relative abundance per sample location and lactation stage are shown and significant differences highlighted (*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

Upon bulk storage there is a shift towards a psychrotolerant-dominant, processing facility-selected, community that is dominated by *Pseudomonas*, *Acinetobacter* and *Lactococcus*. This is consistent with the US study of Kable *et al*. (Kable *et al.*, 2016), which noted that large scale silo storage led to the convergence of the microbiota to one of two community types. The subsequent fate of these bacteria following further product processing in our study showed that the relative abundance of these dominant psychrotrophic bacteria diminished following pasteurisation. This was particularly apparent in mid lactation samples, possibly due to a number of reasons not limited to the lactation stage and associated warmer weather, with an increase in thermophilic bacteria during the processing of these samples, with colder weather and lower production rates associated with late lactation, psychrotrophic dominant, samples. Continued processing of mid lactation samples to a skimmed milk powder had a considerable impact on the microbiota. The resulting powder was dominated by the thermophiles *T. thermophilus* and *Geobacillus sp.*. This dominance may be due to a number of factors including, but not limited to, contamination from within the processing facility (Cho *et al.*, 2018b; Miller *et al.*, 2015), and/or their enrichment due to high temperature treatment, as well at the fact that the first sample was collected during a warmer time of year and when powder production was at its peak, with equipment running for longer periods with higher throughput. The presence of *T. thermophilus* is notable as it has previously been detected in environmental samples from cheese manufacturing facility and has been shown to the pinking defect phenomenon in continental-type cheeses (Quigley *et al.*, 2016). The ability to identify and remove such undesirable

organisms using a HTS incorporated screening regime may provide important economic benefits for food companies downstream.

In contrast to mid lactation samples, the microbiota of late lactation samples did not change as considerably throughout processing. It is not clear if this was due to seasonality or potential process related factors such as different or differences in cleaning practices used at the later sampling time point. Overall significant differences between samples was accounted for by both source of the sample and lactation period. Ultimately, as this was the first such comprehensive study of its kind, further investigation is needed to determine how or why thermophilic species were selected for in one processing run on one day in a mid lactation period, and not selected for in a processing run on one day in a late lactation period. An increased understanding of why and how these differences occurred will undoubtedly aid dairy processors globally to implement effective measures and decrease undesirable microorganisms, while increasing food safety and quality.

Although no previously undetected or unexpected taxa were identified a number of potentially pathogenic bacteria were detected in the samples. One mid lactation bulk tank sample in isolation had a high (> 5%) relative abundance of *Yersinia*, however this accounted for < 0.4% in all processed samples, suggesting its elimination by heat treatments and processing. A number of *Clostridium sensu stricto* subgroups were determined to be present at significantly higher relative abundance in late lactation, compared to mid lactation bulk tanks, however the relative abundance of these also decreased following transport and processing. SUPER-FOCUS functional gene classification noted a higher relative abundance of

sporulation and dormancy associated functional gene groups in mid lactation skimmed milk powder samples, containing thermophilic sporeformers *Geobacillus sp*., than all other samples. Although *Geobacillus sp*. are not pathogenic they are extremely difficult to eliminate from the processing environment due to their heat tolerance, ability to form resistant spores and biofilms which can harbour pathogenic species (Burgess, Flint and Lindsay, 2014; Gopal *et al.*, 2015). *A*. *baumannii* was determined by shotgun metagenomic sequencing to be the dominant species of *Acinetobacter*. Although traditionally associated with opportunistic infection, drug resistance and nosocomial infection, *A*. *baumannii* has been detected in animal products and shown to have different epidemiological characteristics to those strains that cause nosocomial infections (Hamouda *et al.*, 2011). It has also been detected in bulk tank milks and dairy powders, leading to the Food and Agriculture Organization of the United Nations (FAO) to classify it as a "Category B organisms – causality plausible, but not yet demonstrated" with respect to causing infant illness from powdered infant formula (Cho *et al.*, 2018a). The low number of shotgun metagenomic reads did not allow accurate strain level identification, or identification of specific virulence associated genes to facilitate a more in-depth investigation. Despite this, SUPER-FOCUS functional gene classification did show higher relative abundance of virulence associated functional gene groups in raw milk samples in tankers and whole milk silos, compared to heat processed samples in skimmed milk silos and skimmed milk powders. Overall, there is a general pattern of a greater relative abundance of potentially pathogenic genera and species in raw milk samples, and increased relative abundance of spore-forming species present in mid lactation skimmed milk powders. However, it should

be emphasised that relative abundances are reported throughout this manuscript, further analysis would be needed to confirm if there were increases and decreases in absolute abundances. It should be noted that the relative abundance results obtained do largely agree with corresponding culture analysis (Paludetti *et al.*, 2019).

While this study is more specifically focussed on highlighting the merits of using HTS to monitor processing induced-changes in dairy microbiota in general and the fact that these changes can vary, it will be interesting in the future to apply this technology more specifically to elucidate the basis for differences in dairy powder composition. Species level composition analysis from shotgun metagenomic sequencing enables a more in-depth analysis than previously possible with 16S rRNA gene amplicon sequencing. Further development of this method opens the possibility of routine microbiology testing, improved detection of sources of contamination, tracking of microorganisms throughout the food chain and, in general, enhancing the ability of processors to make informed decisions to reduce risk and waste. There are a number of ways in which the approach taken in this study can be improved upon for commercial applications, such as combining with quantitative approaches to get absolute numbers. Untargeted shotgun analysis can be subject to host DNA contamination (Sharpton, 2014), and dairy samples can have particularly high levels of host DNA as somatic cell counts can be high in raw dairy. This can result in decreased yields of microbial DNA sequence, thereby limiting the number of reads available for comprehensive microbial analysis (Supplemental Figure 4.4). Removal of host DNA prior to sequencing is possible by

utilising microbiome enrichment kits (Feehery *et al.*, 2013) and can be considered if performing similar analysis in the future. However, this approach increases the overall cost. Another approach could be to enrich for specific subpopulations prior to DNA extraction (McHugh *et al.*, 2018). Developments such as these, combined with advances in portable sequencing technologies and further advances in the speed and accuracy of *in silico* tools, has the potential to greatly assist decision making in this and other food chains.

4.6 Conclusions

In conclusion, this study provides a detailed insight into the changes in the microbiota of dairy samples throughout a milk powder manufacturing process, on distinct sampling days. A notable change was observed upon large volume pooling in the processing facility, which resulted in the dairy microbiota becoming dominated by psychrotolerant, spoilage-associated bacteria. Also of note were the impacts of processing and the processing facility on the microbiota. A pattern of particular note is the potential proliferation of low levels of thermophilic bacteria present in raw ingredients, or within the processing facility, can potentially proliferate in the absence of competitors during and following processing and dominate the processed dairy product. With routine implementation of these methods, an understanding of the reasons that lead to different species being dominant in final product can be determined and lead to informed decisions regarding product fate, in turn leading to increased food safety, reduced risk and reduced economic losses.

4.7 Declarations

4.7.1 Availability of data

Sequence data has been deposited in the European Nucleotide Archive (ENA) under the study accession number PRJEB31110

4.7.2 Acknowledgements including funding source

This Research was funded by Department of Agriculture, Food and the Marine (DAFM) under the FIRM project 'SACCP', reference number 14/F/883. Research in the Cotter laboratory is also funded by Science Foundation Ireland (SFI) under Grant Numbers SFI/12/RC/2273 (APC Microbiome Ireland) and SFI/16/RC/3835 (Vistamilk) and by the European Commission under the Horizon 2020 programme under grant number 818368 (MASTER).

4.8 References

Alessandria, V., Ferrocino, I., De Filippis, F., Fontana, M., Rantsiou, K., Ercolini, D. and Cocolin, L. (2016) 'Microbiota of an Italian Grana-Like Cheese during Manufacture and Ripening, Unraveled by 16S rRNA-Based Approaches', *Applied and environmental microbiology,* 82(13), pp. 3988-3995. DOI: 10.1128/AEM.00999-16

Altschul, S., Gish, W., Miller, W., Myers, E. and Lipman, D. (1990) 'Basic local alignment search tool', *J Mol Biol,* 215. DOI: 10.1016/s0022-2836(05)80360-2

Burgess, S. A., Flint, S. H. and Lindsay, D. (2014) 'Characterization of thermophilic bacilli from a milk powder processing plant', *J Appl Microbiol,* 116(2), pp. 350-9. DOI: 10.1111/jam.12366

Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L. and Knight, R. (2010a) 'PyNAST: a flexible tool for aligning sequences to a template alignment', *Bioinformatics,* 26(2), pp. 266-267. DOI: 10.1093/bioinformatics/btp636

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. and Knight, R. (2010b) 'QIIME allows analysis of high-throughput community sequencing data', *Nature Methods,* 7, pp. 335. DOI: 10.1038/nmeth.f.303

Cho, G.-S., Li, B., Rostalsky, A., Fiedler, G., Rösch, N., Igbinosa, E., Kabisch, J., Bockelmann, W., Hammer, P., Huys, G. and Franz, C. M. A. P. (2018a) 'Diversity and Antibiotic Susceptibility of Acinetobacter Strains From Milk Powder Produced in Germany', *Frontiers in microbiology,* 9, pp. 536-536. DOI: 10.3389/fmicb.2018.00536

Cho, T. J., Kim, H. W., Kim, N. H., Park, S. M., Kwon, J. I., Kim, Y. J., Lee, K. W. and Rhee, M. S. (2018b) 'New insights into the thermophilic spore-formers in powdered infant formula: Implications of changes in microbial composition during manufacture', *Food Control,* 92, pp. 464-470. DOI: 10.1016/j.foodcont.2018.05.036

Doyle, C. J., Gleeson, D., O'Toole, P. W. and Cotter, P. D. (2017a) 'High-throughput metataxonomic characterization of the raw milk microbiota identifies changes reflecting lactation stage and storage conditions', *Int J Food Microbiol,* 255, pp. 1-6. DOI: 10.1016/j.ijfoodmicro.2017.05.019

Doyle, C. J., Gleeson, D., O'Toole, P. W. and Cotter, P. D. (2017b) 'Impacts of seasonal housing and teat preparation on raw milk microbiota: a high-throughput sequencing study', *Appl Environ Microbiol,* 83(2). DOI: 10.1128/AEM.02694-16

Doyle, C. J., O'Toole, P. W. and Cotter, P. D. (2018) 'Genomic Characterization of Sulphite Reducing Bacteria Isolated From the Dairy Production Chain', *Frontiers in Microbiology,* 9. DOI: 10.3389/fmicb.2018.01507

Edgar, R. C. (2010) 'Search and clustering orders of magnitude faster than BLAST', *Bioinformatics,* 26(19), pp. 2460-2461. DOI: 10.1093/bioinformatics/btq461

Feehery, G. R., Yigit, E., Oyola, S. O., Langhorst, B. W., Schmidt, V. T., Stewart, F. J., Dimalanta, E. T., Amaral-Zettler, L. A., Davis, T., Quail, M. A. and Pradhan, S. (2013) 'A method for selectively enriching microbial DNA from contaminating vertebrate host DNA', *PLoS ONE,* 8(10), pp. e76096. DOI: 10.1371/journal.pone.0076096

Fretin, M., Martin, B., Rifa, E., Isabelle, V. M., Pomies, D., Ferlay, A., Montel, M. C. and Delbes, C. (2018) 'Bacterial community assembly from cow teat skin to ripened cheeses is influenced by grazing systems', *Sci Rep,* 8(1), pp. 200. DOI: 10.1038/s41598-017-18447-y

Fricker, M., Reissbrodt, R. and Ehling-Schulz, M. (2008) 'Evaluation of standard and new chromogenic selective plating media for isolation and identification of Bacillus cereus', *Int J Food Microbiol,* 121(1), pp. 27-34. DOI: 10.1016/j.ijfoodmicro.2007.10.012

Gopal, N., Hill, C., Ross, P. R., Beresford, T. P., Fenelon, M. A. and Cotter, P. D. (2015) 'The prevalence and control of *Bacillus* and related spore-forming bacteria in the dairy industry', *Front Microbiol,* 6, pp. 1418. DOI: 10.3389/fmicb.2015.01418

Hamouda, A., Findlay, J., Al Hassan, L. and Amyes, S. G. B. (2011) 'Epidemiology of Acinetobacter baumannii of animal origin', *International Journal of Antimicrobial Agents,* 38(4), pp. 314-318. DOI: 10.1016/j.ijantimicag.2011.06.007

Kable, M. E., Srisengfa, Y., Laird, M., Zaragoza, J., McLeod, J., Heidenreich, J. and Marco, M. L. (2016) 'The Core and Seasonal Microbiota of Raw Bovine Milk in Tanker Trucks and the Impact of Transfer to a Milk Processing Facility', *MBio,* 7(4). DOI: 10.1128/mBio.00836-16

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) 'The sequence alignment/map format and SAMtools', *Bioinformatics,* 25(16), pp. 2078-9. DOI: 10.1093/bioinformatics/btp352

Li, N., Wang, Y., You, C., Ren, J., Chen, W., Zheng, H. and Liu, Z. (2018) 'Variation in Raw Milk Microbiota Throughout 12 Months and the Impact of Weather Conditions', *Sci Rep,* 8(1), pp. 2371. DOI: 10.1038/s41598-018-20862-8

Magoc, T. and Salzberg, S. L. (2011) 'FLASH: fast length adjustment of short reads to improve genome assemblies', *Bioinformatics,* 27(21), pp. 2957-63. DOI: 10.1093/bioinformatics/btr507

McHugh, A. J., Feehily, C., Tobin, J. T., Fenelon, M. A., Hill, C. and Cotter, P. D. (2018) 'Mesophilic Sporeformers Identified in Whey Powder by Using Shotgun Metagenomic Sequencing', *Applied and Environmental Microbiology,* 84(20). DOI: 10.1128/AEM.01305-18

McMurdie, P. J. and Holmes, S. (2013) 'phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data', *PLOS ONE,* 8(4), pp. e61217. DOI: 10.1371/journal.pone.0061217

Miller, R. A., Kent, D. J., Watterson, M. J., Boor, K. J., Martin, N. H. and Wiedmann, M. (2015) 'Spore populations among bulk tank raw milk and dairy powders are significantly different', *J Dairy Sci,* 98(12), pp. 8492-504. DOI: 10.3168/jds.2015-9943

Paludetti, L. F., Kelly, A. L., O'Brien, B., Jordan, K. and Gleeson, D. (2019) 'Microbiological quality of milk from farms to milk powder manufacture: an industrial case study', *J Dairy Res*, pp. 1-6. DOI: 10.1017/S0022029919000347

Parente, E., Cocolin, L., De Filippis, F., Zotta, T., Ferrocino, I., O'Sullivan, O., Neviani, E., De Angelis, M., Cotter, P. D. and Ercolini, D. (2016) 'FoodMicrobionet: A database for the visualisation and exploration of food bacterial communities based on network analysis', *Int J Food Microbiol,* 219, pp. 28-37. DOI: 10.1016/j.ijfoodmicro.2015.12.001

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F. O. (2013) 'The SILVA ribosomal RNA gene database project: improved data processing and web-based tools', *Nucleic Acids Research,* 41(D1), pp. D590-D596. DOI: 10.1093/nar/gks1219

Quigley, L., McCarthy, R., O'Sullivan, O., Beresford, T. P., Fitzgerald, G. F., Ross, R. P., Stanton, C. and Cotter, P. D. (2013) 'The microbial content of raw and pasteurized cow milk as determined by molecular approaches', *J Dairy Sci,* 96(8), pp. 4928-37. DOI: 10.3168/jds.2013-6688

Quigley, L., O'Sullivan, D. J., Daly, D., O'Sullivan, O., Burdikova, Z., Vana, R., Beresford, T. P., Ross, R. P., Fitzgerald, G. F., McSweeney, P. L., Giblin, L., Sheehan, J. J. and Cotter, P. D. (2016) 'Thermus and the pink discoloration defect in cheese', *mSystems,* 1(3). DOI: 10.1128/mSystems.00023-16

Quijada, N. M., Mann, E., Wagner, M., Rodriguez-Lazaro, D., Hernandez, M. and Schmitz-Esser, S. (2018) 'Autochthonous facility-specific microbiota dominates

washed-rind Austrian hard cheese surfaces and its production environment', *Int J Food Microbiol,* 267, pp. 54-61. DOI: 10.1016/j.ijfoodmicro.2017.12.025

Rodrigues, M. X., Lima, S. F., Canniatti-Brazaca, S. G. and Bicalho, R. C. (2017) 'The microbiome of bulk tank milk: Characterization and associations with somatic cell count and bacterial count', *J Dairy Sci*. DOI: 10.3168/jds.2016-11540

Sadiq, F. A., Flint, S. and He, G. (2018) 'Microbiota of milk powders and the heat resistance and spoilage potential of aerobic spore-forming bacteria', *International Dairy Journal,* 85, pp. 159-168. DOI: 10.1016/j.idairyj.2018.06.003

Sharpton, T. J. (2014) 'An introduction to the analysis of shotgun metagenomic data', *Front Plant Sci,* 5, pp. 209. DOI: 10.3389/fpls.2014.00209

Silva, G. G., Green, K. T., Dutilh, B. E. and Edwards, R. A. (2016) 'SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data', *Bioinformatics,* 32(3), pp. 354-61. DOI: 10.1093/bioinformatics/btv584

Stellato, G., Utter, D. R., Voorhis, A., De Angelis, M., Eren, A. M. and Ercolini, D. (2017) 'A Few Pseudomonas Oligotypes Dominate in the Meat and Dairy Processing Environment', *Frontiers in microbiology,* 8, pp. 264-264. DOI: 10.3389/fmicb.2017.00264

Walsh, A. M., Crispie, F., Daari, K., O'Sullivan, O., Martin, J. C., Arthur, C. T., Claesson, M. J., Scott, K. P. and Cotter, P. D. (2017) 'Strain-level metagenomic analysis of the fermented dairy beverage nunu highlights potential food safety risks', *Applied and Environmental Microbiology*. DOI: 10.1128/aem.01144-17

Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis.* Springer Publishing Company, Incorporated. DOI: 10.1007/978-0-387-98141-3

Wolfe, B. E., Button, J. E., Santarelli, M. and Dutton, R. J. (2014) 'Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity', *Cell,* 158(2), pp. 422-433. DOI: 10.1016/j.cell.2014.05.041

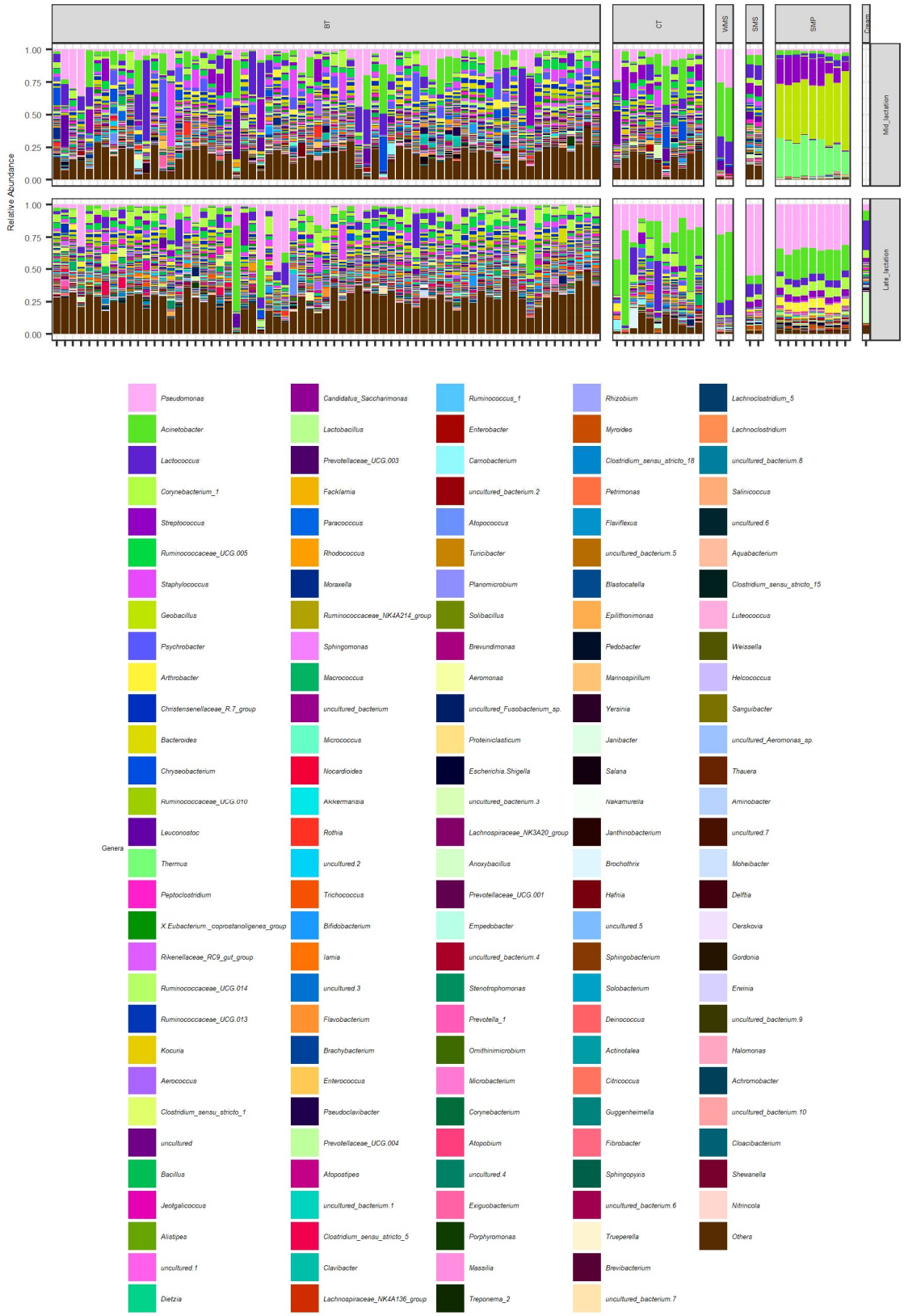
Wood, D. E. and Salzberg, S. L. (2014) 'Kraken: ultrafast metagenomic sequence classification using exact alignments', *Genome Biology,* 15(3), pp. R46. DOI: 10.1186/gb-2014-15-3-r46

Wu, S., Jiang, Y., Lou, B., Feng, J., Zhou, Y., Guo, L., Forsythe, S. J. and Man, C. (2018) 'Microbial community structure and distribution in the air of a powdered infant formula factory based on cultivation and high-throughput sequence methods', *J Dairy Sci*. DOI: 10.3168/jds.2017-13968

Yeluri Jonnala, B. R., McSweeney, P. L. H., Sheehan, J. J. and Cotter, P. D. (2018) 'Sequencing of the Cheese Microbiome and Its Relevance to Industry', *Front Microbiol,* 9, pp. 1020. DOI: 10.3389/fmicb.2018.01020

Zhao, Y., Kumar, M., Caspers, M. P. M., Nierop Groot, M. N., van der Vossen, J. and Abee, T. (2018) 'Short communication: Growth of dairy isolates of Geobacillus thermoglucosidans in skim milk depends on lactose degradation products supplied by Anoxybacillus flavithermus as secondary species', *J Dairy Sci*. DOI: 10.3168/jds.2017-13372

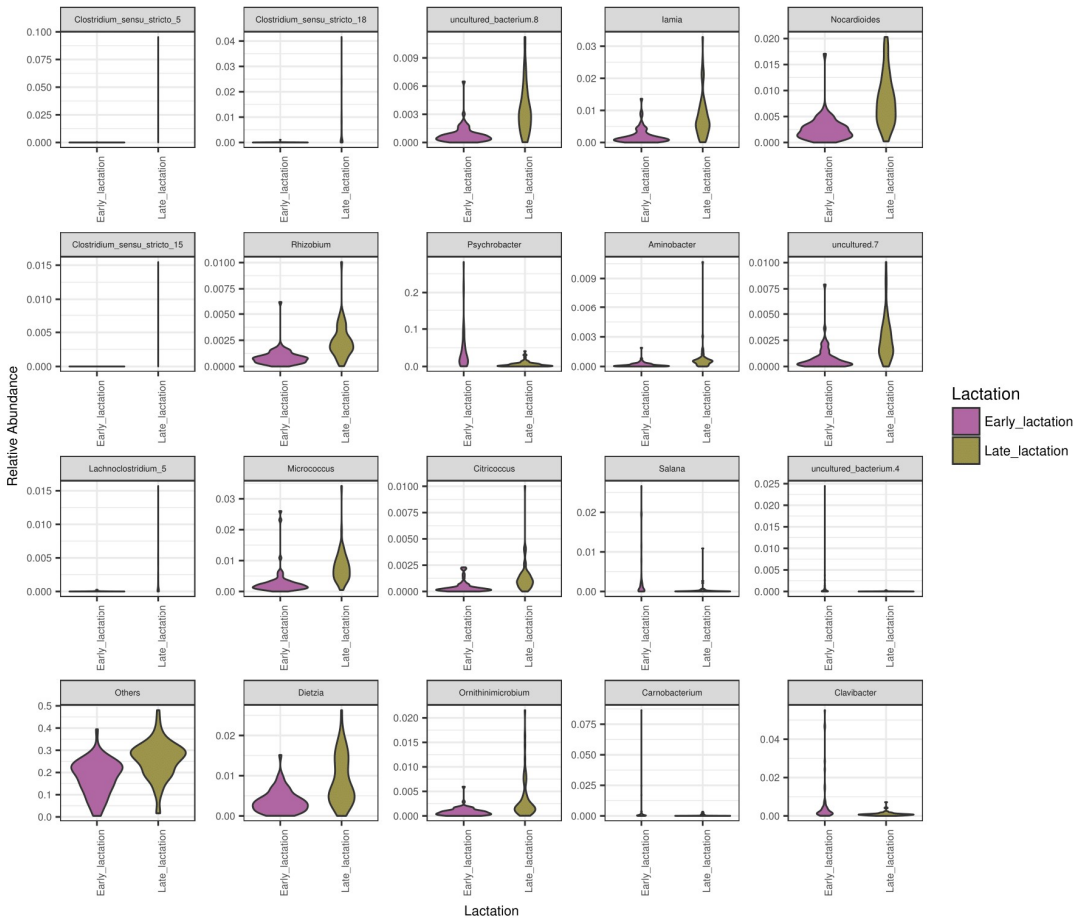
4.9 Supplemental Figures

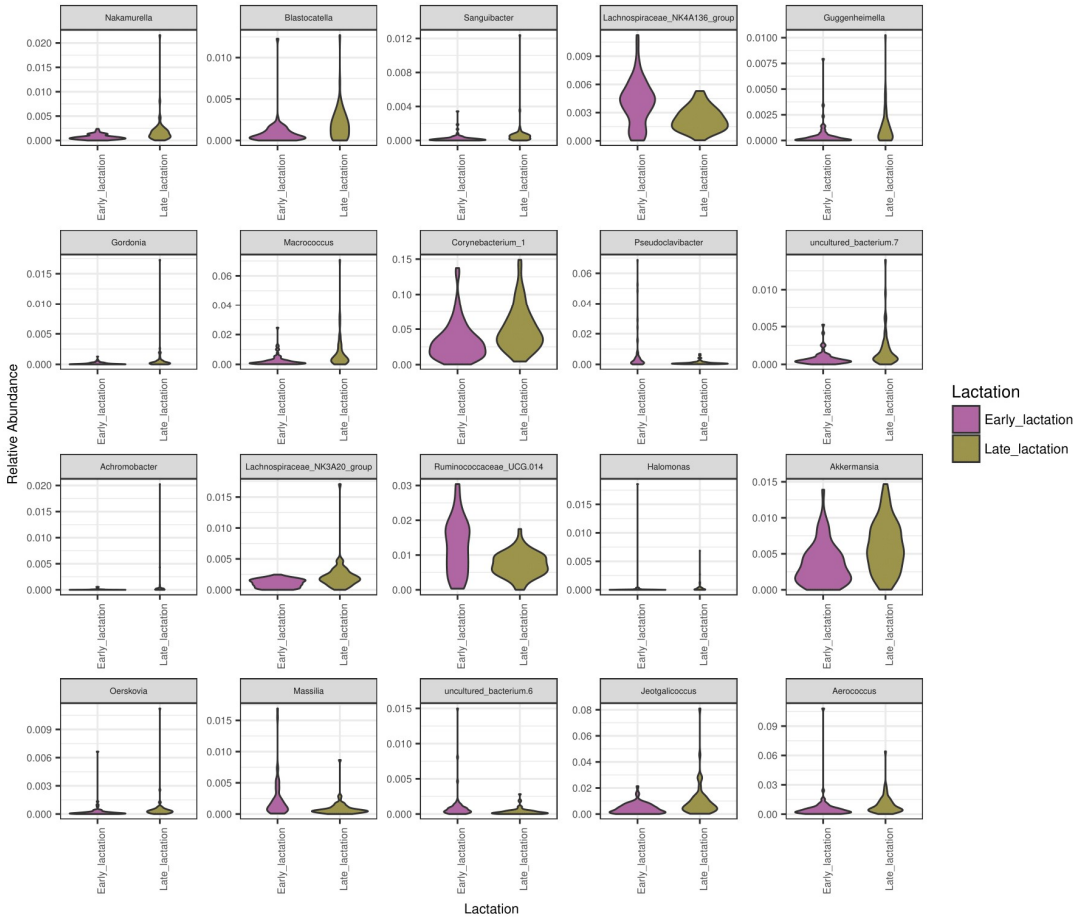


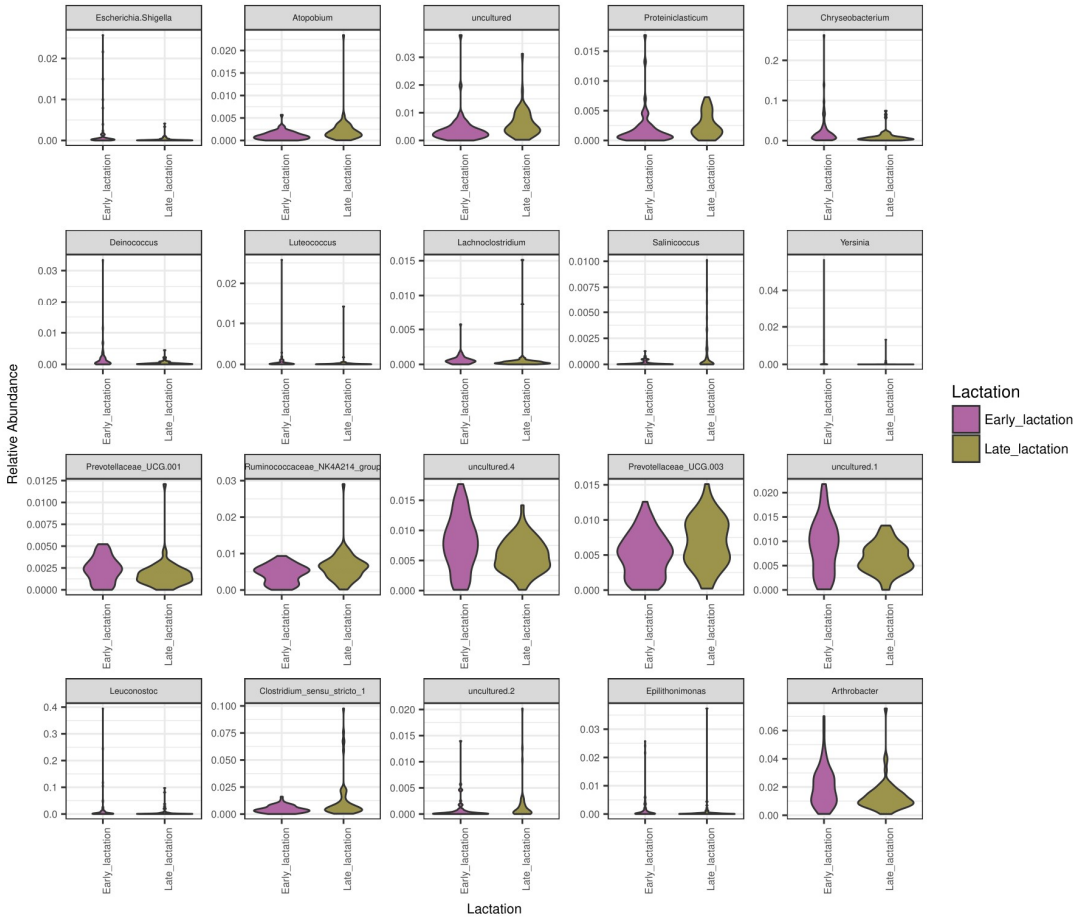
Supplemental Figure 4.1. Relative abundance of genera present > 1% relative abundance per sample.

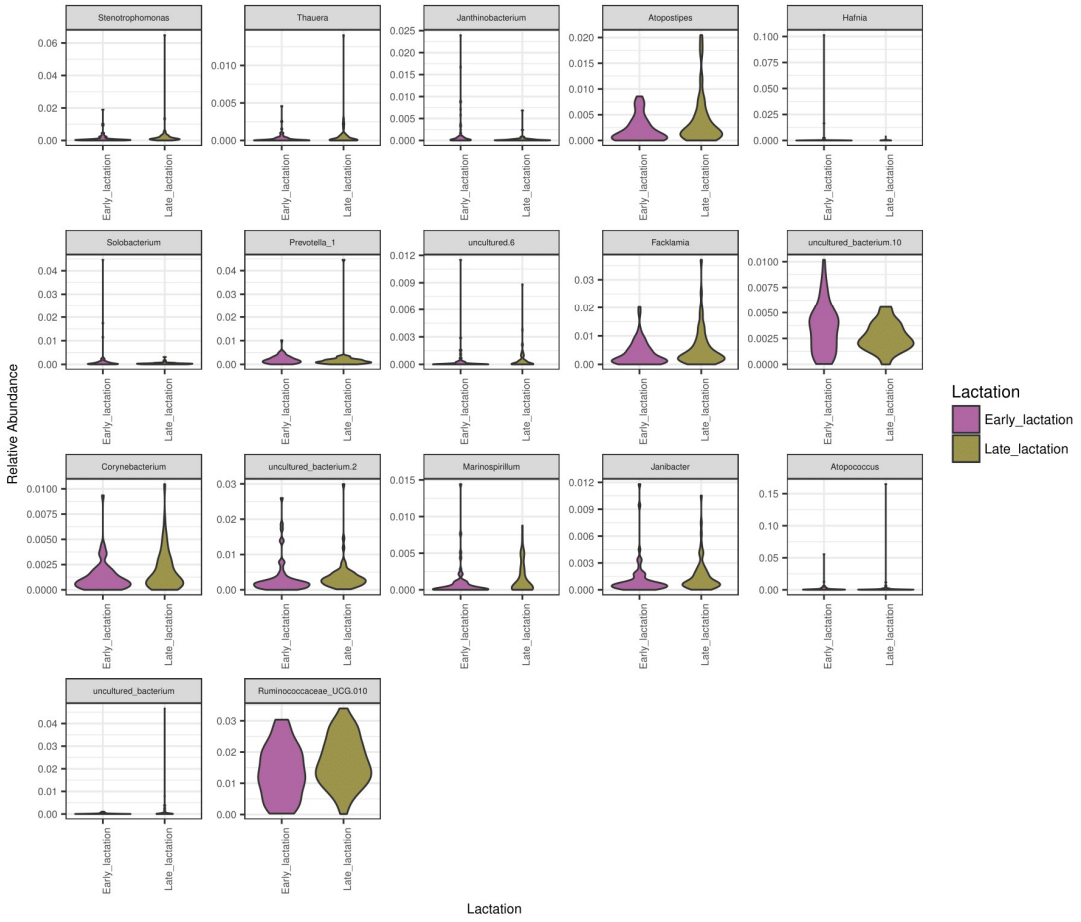
[Supplemental Figure 4.1 continued]

Genera level taxonomic classification from 16S rRNA gene amplicon sequence data for mid and late lactation samples from bulk tanks (BT), collection tankers (CT), whole milk silos (WMS), skimmed milk silos (SMS), skimmed milk powder (SMP) and cream (late lactation only) samples. Genera shown are present in > 1% relative abundance in at least one sample.


Clostridium_sensu_stricto_5
Clostridium_sensu_stricto_18
uncultured_bacterium.8
Iamia
Nocardioides
Clostridium_sensu_stricto_15
Rhizobium
Psychrobacter
Aminobacter
uncultured.7
Lachnoclostridium_5
Micrococcus
Citricoccus
Salana
uncultured_bacterium.4
Others
Dietzia
Ornithinimicrobium
Carnobacterium
Clavibacter
Early_lactation
Late_lactation
Relative Abundance
Lactation
Lactation
Early_lactation
Late_lactation

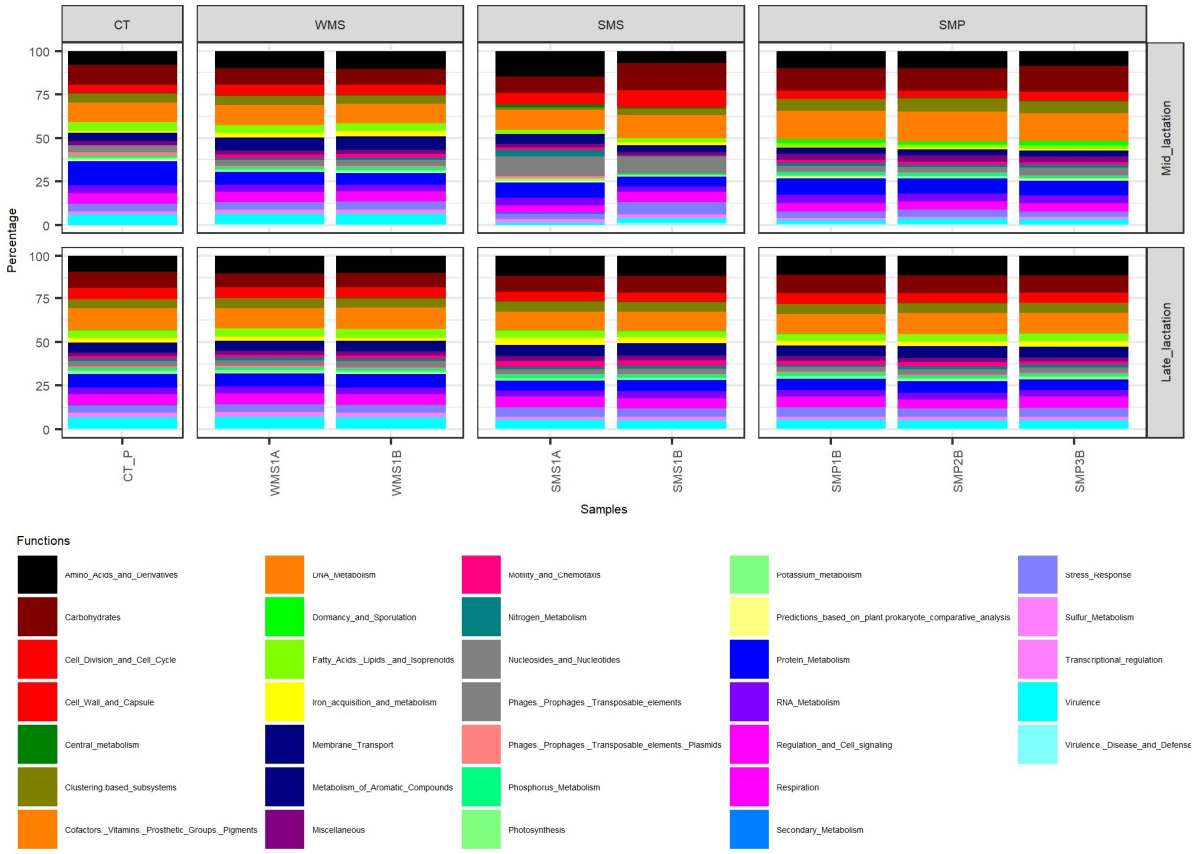

Nakamurella
Blastocatella
Sanguibacter
Lachnospiraceae_NK4A136_group
Guggenheimella
Gordonia
Macrococcus
Corynebacterium_1
Pseudoclavibacter
uncultured_bacterium.7
Achromobacter
Lachnospiraceae_NK3A20_group
Ruminococcaceae_UCG.014
Halomonas
Akkermansia
Oerskovia
Massilia
uncultured_bacterium.6
Jeotgalicoccus
Aerococcus
Early_lactation
Late_lactation
Relative Abundance
Lactation
Lactation
Early_lactation
Late_lactation


Escherichia.Shigella
Atopobium
uncultured
Proteiniclasticum
Chryseobacterium
Deinococcus
Luteococcus
Lachnoclostridium
Salinicoccus
Yersinia
Prevotellaceae_UCG.001
Ruminococcaceae_NK4A214_group
uncultured.4
Prevotellaceae_UCG.003
uncultured.1
Leuconostoc
Clostridium_sensu_stricto_1
uncultured.2
Epilithonimonas
Arthrobacter
Early_lactation
Late_lactation
Relative Abundance
Lactation
Lactation
Early_lactation
Late_lactation



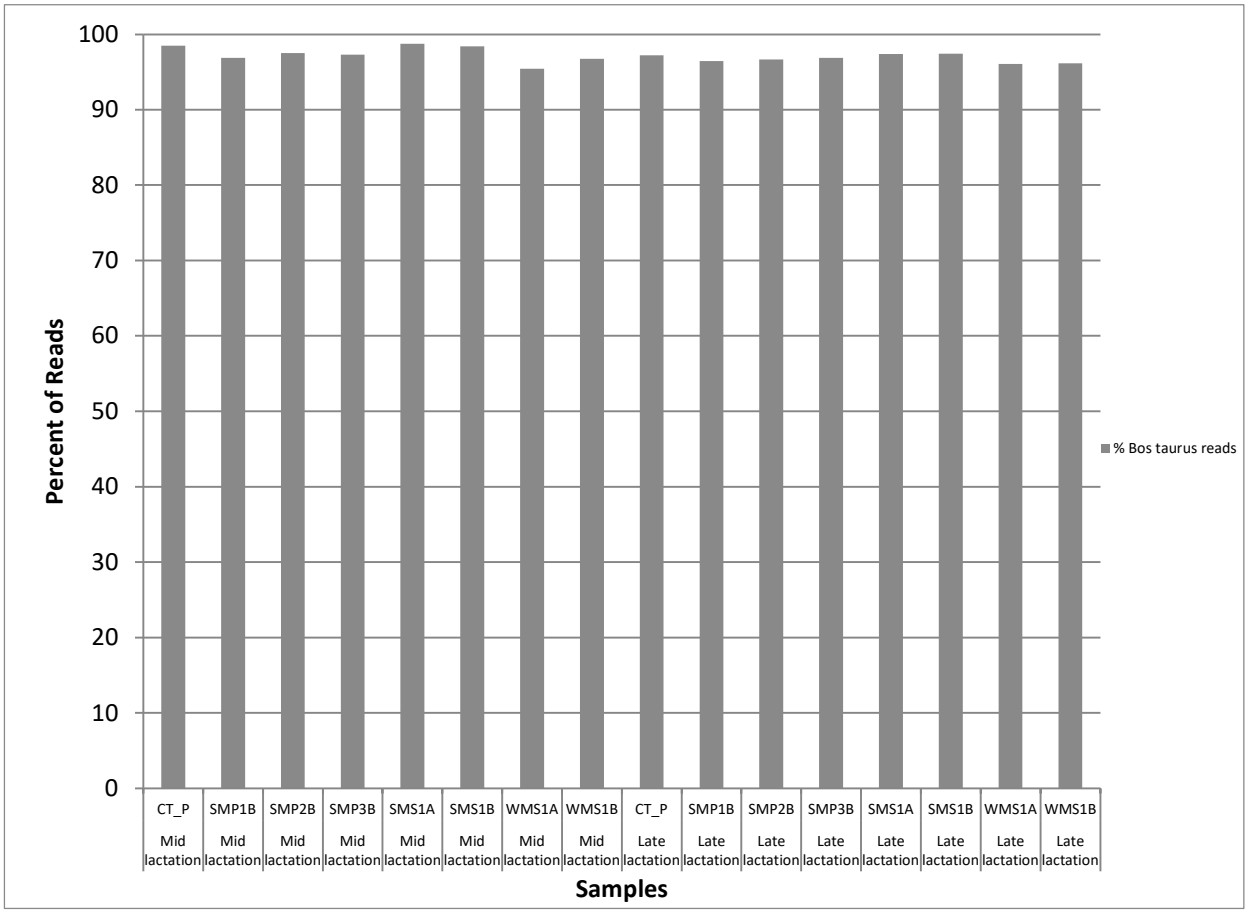
Supplemental Figure 4.2. Significantly differential relative abundance of taxa constituting at least 1% relative abundance of at least one sample between mid and late lactation bulk tanks.

Seventy-seven significantly different taxa > 1% relative abundance between mid and late lactation.



Supplemental Figure 4.3. SUPER-FOCUS L1 functional analysis.

SUPER-FOCUS L1 functional analysis on a subset of samples from the mid and late lactation process, including a pooled representative collection tanker sample (CT_P), whole milk silos (WMS), skimmed milk silos (SMS) and a subset of skimmed milk powder samples (SMP) from each lactation stage.



Supplemental Figure 4.4. Percentage of *Bos taurus* aligned reads per shotgun metagenomic sample.

A high percent of *Bos taurus* aligned reads per shotgun metagenomic sample.

Supplemental Table 4.1. Table summarising the genera with the highest relative abundance per group.

		Mid lactation					**Late lactation**					
Genera	**Stat**	**BT**	**CT**	**WMS**	**SMS**	**SMP**	**BT**	**CT**	**WMS**	**SMS**	**SMP**	**Cream**
Pseudomonas	Mean (± SD)	***6.58E-02 (± 9.01E-02)***	***7.24E-02 (± 6.29E-02)***	***2.73E-01 (± 2.78E-02)***	4.07E-02 (± 4.52E-04)	1.49E-02 (± 9.27E-03)	***6.59E-02 (± 1.03E-01)***	***2.03E-01 (± 9.77E-02)***	***2.19E-01 (± 1.23E-02)***	***5.49E-01 (± 3.54E-03)***	***3.43E-01 (± 2.07E-02)***	4.49E-02
Pseudomonas	Min ; Max	***3.41E-03 ; 4.35E-01***	***1.56E-02 ; 2.32E-01***	***2.53E-01 ; 2.92E-01***	4.03E-02 ; 4.10E-02	7.34E-03 ; 3.24E-02	***2.35E-03 ; 5.19E-01***	***1.01E-01 ; 4.19E-01***	***2.11E-01 ; 2.28E-01***	***5.47E-01 ; 5.52E-01***	***3.09E-01 ; 3.86E-01***	4.49E-02
Acinetobacter	Mean (± SD)	***6.23E-02 (± 8.42E-02)***	***9.65E-02 (± 1.08E-01)***	***4.13E-01 (± 1.15E-04)***	***7.73E-02 (± 3.99E-03)***	2.95E-02 (± 9.13E-03)	***4.14E-02 (± 9.09E-02)***	***2.54E-01 (± 2.17E-01)***	***5.31E-01 (± 3.61E-03)***	***6.68E-02 (± 1.24E-03)***	***2.14E-01 (± 2.62E-02)***	***7.69E-02***
Acinetobacter	Min ; Max	***3.19E-03 ; 3.77E-01***	***1.36E-02 ; 3.04E-01***	***4.13E-01 ; 4.13E-01***	***7.44E-02 ; 8.01E-02***	1.71E-02 ; 4.80E-02	***3.08E-03 ; 6.60E-01***	***3.19E-02 ; 7.40E-01***	***5.28E-01 ; 5.33E-01***	***6.60E-02 ; 6.77E-02***	***1.83E-01 ; 2.43E-01***	***7.69E-02***
Lactococcus	Mean (± SD)	***5.98E-02 (± 1.05E-01)***	***5.21E-02 (± 4.88E-02)***	***1.62E-01 (± 9.36E-03)***	***1.08E-01 (± 8.24E-03)***	***3.05E-02 (± 1.26E-02)***	***3.50E-02 (± 4.87E-02)***	***7.05E-02 (± 1.23E-01)***	***1.05E-01 (± 1.37E-02)***	***8.92E-02 (± 9.63E-03)***	***5.05E-02 (± 7.02E-03)***	***2.27E-01***
Lactococcus	Min ; Max	***2.69E-04 ; 6.19E-01***	***5.10E-03 ; 1.36E-01***	***1.55E-01 ; 1.68E-01***	***1.02E-01 ; 1.14E-01***	***1.45E-02 ; 5.01E-02***	***5.58E-04 ; 2.62E-01***	***2.57E-03 ; 3.93E-01***	***9.55E-02 ; 1.15E-01***	***8.24E-02 ; 9.60E-02***	***3.56E-02 ; 5.72E-02***	***2.27E-01***
Corynebacterium 1	Mean (± SD)	***3.30E-02 (± 2.70E-02)***	***3.28E-02 (± 1.86E-02)***	1.65E-03 (± 4.74E-04)	4.02E-02 (± 6.95E-04)	5.05E-03 (± 3.03E-03)	***5.06E-02 (± 3.09E-02)***	***5.59E-02 (± 3.82E-02)***	1.68E-02 (± 8.05E-04)	3.65E-02 (± 6.81E-04)	***7.85E-02 (± 8.42E-03)***	***5.96E-02***
Corynebacterium 1	Min ; Max	***3.94E-04 ; 1.37E-01***	***2.98E-03 ; 5.67E-02***	1.31E-03 ; 1.99E-03	3.97E-02 ; 4.07E-02	2.01E-03 ; 1.05E-02	***4.44E-03 ; 1.49E-01***	***4.97E-03 ; 1.31E-01***	1.62E-02 ; 1.73E-02	3.60E-02 ; 3.70E-02	***6.88E-02 ; 9.36E-02***	***5.96E-02***
Streptococcus	Mean (± SD)	***3.49E-02 (± 6.91E-02)***	***6.94E-02 (± 7.63E-02)***	2.71E-03 (± 1.32E-03)	***8.53E-02 (± 1.43E-04)***	***1.60E-01 (± 5.70E-02)***	***1.45E-02 (± 3.15E-02)***	***1.95E-02 (± 2.56E-02)***	8.61E-03 (± 2.99E-04)	***5.36E-02 (± 8.05E-03)***	***4.60E-02 (± 5.80E-03)***	3.33E-02
Streptococcus	Min ; Max	***8.40E-04 ; 3.61E-01***	***5.09E-03 ; 2.56E-01***	1.78E-03 ; 3.65E-03	***8.52E-02 ; 8.54E-02***	***6.80E-02 ; 2.13E-01***	***8.30E-04 ; 2.51E-01***	***8.62E-04 ; 8.63E-02***	8.40E-03 ; 8.82E-03	***4.79E-02 ; 5.93E-02***	***3.82E-02 ; 5.54E-02***	3.33E-02

Ruminococcaceae UCG.005	Mean (± SD)	***3.83E-02 (± 2.10E-02)***	***3.92E-02 (± 1.42E-02)***	1.64E-03 (± 6.45E-04)	3.31E-02 (± 1.81E-03)	3.88E-04 (± 1.31E-04)	***3.78E-02 (± 1.58E-02)***	1.48E-02 (± 1.07E-02)	3.13E-03 (± 1.68E-04)	4.21E-03 (± 5.13E-04)	2.19E-03 (± 1.04E-03)	7.57E-03
Ruminococcaceae UCG.005	Min ; Max	***5.72E-04 ; 8.64E-02***	***5.72E-03 ; 5.20E-02***	1.19E-03 ; 2.10E-03	3.19E-02 ; 3.44E-02	2.39E-04 ; 5.73E-04	***6.42E-04 ; 7.87E-02***	1.09E-03 ; 3.15E-02	3.01E-03 ; 3.25E-03	3.85E-03 ; 4.57E-03	1.42E-03 ; 4.58E-03	7.57E-03
Staphylococcus	Mean (± SD)	***3.82E-02 (± 6.78E-02)***	***3.14E-02 (± 3.82E-02)***	2.06E-03 (± 8.30E-04)	3.25E-02 (± 4.57E-03)	2.37E-03 (± 1.01E-03)	***2.63E-02 (± 3.71E-02)***	6.35E-03 (± 8.45E-03)	4.12E-03 (± 5.86E-04)	3.63E-03 (± 2.26E-04)	7.96E-03 (± 2.11E-03)	1.19E-02
Staphylococcus	Min ; Max	***1.18E-03 ; 4.81E-01***	***3.06E-03 ; 1.35E-01***	1.47E-03 ; 2.64E-03	2.93E-02 ; 3.58E-02	1.35E-03 ; 4.21E-03	***2.43E-03 ; 2.70E-01***	5.24E-04 ; 3.07E-02	3.71E-03 ; 4.54E-03	3.47E-03 ; 3.79E-03	5.07E-03 ; 1.18E-02	1.19E-02
Geobacillus	Mean (± SD)	5.19E-04 (± 9.64E-04)	3.35E-03 (± 4.86E-03)	1.87E-04 (± 2.52E-04)	2.06E-03 (± 2.37E-03)	***4.61E-01 (± 7.12E-02)***	2.78E-04 (± 4.12E-04)	1.04E-04 (± 8.51E-05)	4.41E-05 (± 2.26E-05)	5.26E-05 (± 7.44E-05)	1.38E-04 (± 1.01E-04)	5.12E-05
Geobacillus	Min ; Max	0.00E+00 ; 5.74E-03	1.26E-04 ; 1.44E-02	8.59E-06 ; 3.65E-04	3.78E-04 ; 3.73E-03	***4.04E-01 ; 6.04E-01***	0.00E+00 ; 1.86E-03	3.02E-05 ; 3.18E-04	2.82E-05 ; 6.01E-05	0.00E+00 ; 1.05E-04	6.04E-05 ; 3.72E-04	5.12E-05
Psychrobacter	Mean (± SD)	***4.01E-02 (± 5.15E-02)***	1.33E-02 (± 8.77E-03)	1.07E-03 (± 6.43E-04)	1.23E-02 (± 4.44E-04)	1.16E-03 (± 7.34E-04)	6.01E-03 (± 7.91E-03)	***2.44E-02 (± 2.23E-02)***	4.36E-03 (± 7.58E-05)	4.17E-03 (± 7.56E-04)	2.67E-03 (± 4.61E-04)	4.75E-03
Psychrobacter	Min ; Max	***4.73E-04 ; 2.82E-01***	3.80E-03 ; 3.27E-02	6.19E-04 ; 1.53E-03	1.20E-02 ; 1.26E-02	4.13E-04 ; 2.54E-03	7.05E-06 ; 4.10E-02	***6.77E-04 ; 6.99E-02***	4.31E-03 ; 4.42E-03	3.64E-03 ; 4.71E-03	2.15E-03 ; 3.66E-03	4.75E-03
Arthrobacter	Mean (± SD)	***2.01E-02 (± 1.37E-02)***	9.50E-03 (± 5.49E-03)	1.28E-03 (± 2.74E-04)	6.34E-03 (± 8.55E-04)	4.27E-04 (± 2.16E-04)	***1.33E-02 (± 1.08E-02)***	***1.82E-02 (± 1.86E-02)***	6.07E-03 (± 4.35E-04)	1.39E-02 (± 1.00E-03)	***6.53E-02 (± 6.73E-03)***	7.71E-03
Arthrobacter	Min ; Max	***9.28E-04 ; 7.02E-02***	1.63E-03 ; 1.99E-02	1.09E-03 ; 1.48E-03	5.73E-03 ; 6.94E-03	2.31E-04 ; 8.55E-04	***8.70E-04 ; 7.57E-02***	***1.08E-03 ; 5.94E-02***	5.76E-03 ; 6.38E-03	1.32E-02 ; 1.46E-02	***5.56E-02 ; 7.40E-02***	7.71E-03
Christensenellaceae R.7 group	Mean (± SD)	1.81E-02 (± 1.03E-02)	1.66E-02 (± 7.17E-03)	9.70E-04 (± 3.52E-04)	1.39E-02 (± 1.03E-03)	3.00E-04 (± 1.22E-04)	***2.09E-02 (± 1.03E-02)***	7.19E-03 (± 4.75E-03)	2.08E-03 (± 6.56E-05)	2.51E-03 (± 5.74E-05)	1.61E-03 (± 6.78E-04)	6.28E-03
Christensenellaceae R.7 group	Min ; Max	2.80E-04 ; 3.81E-02	2.28E-03 ; 2.92E-02	7.22E-04 ; 1.22E-03	1.32E-02 ; 1.47E-02	1.16E-04 ; 4.79E-04	***7.85E-04 ; 7.29E-02***	5.73E-04 ; 1.34E-02	2.04E-03 ; 2.13E-03	2.47E-03 ; 2.55E-03	9.46E-04 ; 2.92E-03	6.28E-03
Bacteroides	Mean (± SD)	1.87E-02 (± 1.08E-02)	1.74E-02 (± 7.01E-03)	8.98E-04 (± 2.86E-04)	1.66E-02 (± 4.52E-03)	1.48E-03 (± 9.40E-04)	***1.94E-02 (± 1.46E-02)***	7.95E-03 (± 6.38E-03)	1.43E-03 (± 1.39E-04)	1.79E-03 (± 3.75E-04)	1.77E-03 (± 3.45E-04)	4.20E-03

Bacteroides	Min ; Max	3.43E-04 ; 4.24E-02	6.37E-03 ; 2.61E-02	6.96E-04 ; 1.10E-03	1.34E-02 ; 1.98E-02	3.61E-04 ; 2.94E-03	***3.57E-04 ; 1.14E-01***	5.10E-04 ; 1.77E-02	1.33E-03 ; 1.53E-03	1.52E-03 ; 2.05E-03	1.33E-03 ; 2.51E-03	4.20E-03
Chryseobacterium	Mean (± SD)	***2.37E-02 (± 3.85E-02)***	***3.72E-02 (± 5.86E-02)***	1.52E-02 (± 1.06E-03)	1.71E-02 (± 1.84E-03)	9.86E-05 (± 7.64E-05)	***9.89E-03 (± 1.35E-02)***	5.94E-03 (± 7.18E-03)	1.08E-03 (± 5.01E-04)	4.15E-04 (± 2.11E-05)	1.15E-04 (± 6.35E-05)	5.63E-04
Chryseobacterium	Min ; Max	***8.33E-04 ; 2.63E-01***	***3.91E-03 ; 1.63E-01***	1.45E-02 ; 1.60E-02	1.57E-02 ; 1.84E-02	2.48E-05 ; 2.83E-04	***4.02E-04 ; 7.45E-02***	2.18E-04 ; 2.07E-02	7.21E-04 ; 1.43E-03	4.00E-04 ; 4.30E-04	2.40E-05 ; 2.08E-04	5.63E-04
Leuconostoc	Mean (± SD)	***1.80E-02 (± 5.86E-02)***	***3.68E-02 (± 7.55E-02)***	***6.26E-02 (± 1.11E-02)***	***5.10E-02 (± 9.88E-03)***	4.32E-03 (± 2.93E-03)	***6.00E-03 (± 1.64E-02)***	4.33E-03 (± 4.99E-03)	1.06E-03 (± 3.28E-04)	1.56E-03 (± 5.25E-04)	7.46E-04 (± 1.90E-04)	8.53E-04
Leuconostoc	Min ; Max	***1.27E-05 ; 3.95E-01***	***2.96E-04 ; 2.49E-01***	***5.48E-02 ; 7.05E-02***	***4.40E-02 ; 5.80E-02***	1.29E-03 ; 9.34E-03	***0.00E+00 ; 9.75E-02***	1.92E-04 ; 1.84E-02	8.33E-04 ; 1.30E-03	1.19E-03 ; 1.93E-03	4.83E-04 ; 1.07E-03	8.53E-04
Thermus	Mean (± SD)	1.84E-05 (± 4.16E-05)	9.91E-05 (± 1.80E-04)	5.06E-05 (± 5.94E-05)	1.78E-04 (± 7.71E-05)	***2.54E-01 (± 4.69E-02)***	1.98E-04 (± 5.68E-04)	1.86E-05 (± 1.41E-05)	4.29E-06 (± 6.07E-06)	0.00E+00 (± 0.00E+00)	4.21E-05 (± 4.14E-05)	1.71E-05
Thermus	Min ; Max	0.00E+00 ; 2.17E-04	0.00E+00 ; 5.54E-04	8.59E-06 ; 9.26E-05	1.23E-04 ; 2.32E-04	***1.76E-01 ; 3.12E-01***	0.00E+00 ; 3.64E-03	0.00E+00 ; 3.46E-05	0.00E+00 ; 8.59E-06	0.00E+00 ; 0.00E+00	0.00E+00 ; 1.33E-04	1.71E-05
Peptoclostridium	Mean (± SD)	1.23E-02 (± 6.92E-03)	1.15E-02 (± 4.11E-03)	7.44E-04 (± 5.18E-04)	3.46E-02 (± 8.21E-05)	2.11E-03 (± 7.67E-04)	***1.49E-02 (± 1.20E-02)***	3.63E-03 (± 2.77E-03)	1.12E-03 (± 7.41E-05)	7.61E-03 (± 4.57E-04)	8.37E-03 (± 2.15E-03)	2.02E-02
Peptoclostridium	Min ; Max	2.54E-04 ; 3.16E-02	1.87E-03 ; 1.58E-02	3.78E-04 ; 1.11E-03	3.46E-02 ; 3.47E-02	1.39E-03 ; 3.82E-03	***1.43E-04 ; 6.94E-02***	3.26E-04 ; 8.04E-03	1.06E-03 ; 1.17E-03	7.29E-03 ; 7.94E-03	5.97E-03 ; 1.25E-02	2.02E-02
Kocuria	Mean (± SD)	7.43E-03 (± 7.95E-03)	1.29E-02 (± 1.56E-02)	3.30E-04 (± 7.84E-05)	1.61E-02 (± 4.05E-04)	1.21E-03 (± 6.98E-04)	***8.56E-03 (± 8.75E-03)***	3.82E-03 (± 2.20E-03)	1.44E-03 (± 3.42E-04)	2.53E-03 (± 2.35E-05)	6.03E-03 (± 5.36E-04)	4.92E-03
Kocuria	Min ; Max	5.08E-05 ; 3.12E-02	4.45E-04 ; 4.36E-02	2.75E-04 ; 3.86E-04	1.58E-02 ; 1.64E-02	4.13E-04 ; 2.25E-03	***4.70E-04 ; 5.58E-02***	3.08E-04 ; 6.37E-03	1.19E-03 ; 1.68E-03	2.51E-03 ; 2.54E-03	5.29E-03 ; 6.81E-03	4.92E-03
Aerococcus	Mean (± SD)	***6.18E-03 (± 1.33E-02)***	4.49E-03 (± 2.36E-03)	6.53E-04 (± 3.62E-05)	5.59E-03 (± 2.84E-04)	4.33E-04 (± 3.05E-04)	***9.35E-03 (± 9.57E-03)***	8.52E-03 (± 5.99E-03)	3.81E-03 (± 7.94E-04)	6.83E-03 (± 1.01E-04)	9.40E-03 (± 1.68E-03)	7.23E-03
Aerococcus	Min ; Max	***1.94E-04 ; 1.08E-01***	1.55E-03 ; 8.74E-03	6.27E-04 ; 6.79E-04	5.39E-03 ; 5.79E-03	1.62E-04 ; 1.02E-03	***1.57E-04 ; 6.42E-02***	6.00E-04 ; 2.22E-02	3.25E-03 ; 4.37E-03	6.76E-03 ; 6.90E-03	7.11E-03 ; 1.17E-02	7.23E-03

Clostridium sensu stricto 1	Mean (± SD)	4.56E-03 (± 3.30E-03)	4.47E-03 (± 2.16E-03)	1.96E-04 (± 1.08E-04)	9.02E-03 (± 1.63E-03)	5.68E-04 (± 2.60E-04)	***1.21E-02 (± 1.88E-02)***	1.67E-03 (± 1.32E-03)	3.70E-04 (± 7.43E-05)	1.69E-03 (± 1.46E-04)	1.68E-03 (± 6.40E-04)	5.56E-03
Clostridium sensu stricto 1	Min ; Max	7.62E-05 ; 1.61E-02	1.20E-03 ; 8.90E-03	1.20E-04 ; 2.73E-04	7.87E-03 ; 1.02E-02	2.80E-04 ; 9.73E-04	***3.14E-04 ; 9.77E-02***	1.49E-04 ; 4.13E-03	3.18E-04 ; 4.23E-04	1.59E-03 ; 1.79E-03	1.21E-03 ; 3.20E-03	5.56E-03
Bacillus	Mean (± SD)	***6.46E-03 (± 9.48E-03)***	***1.23E-02 (± 1.75E-02)***	4.66E-04 (± 5.99E-04)	6.39E-03 (± 6.00E-03)	4.49E-04 (± 2.41E-04)	***8.49E-03 (± 1.03E-02)***	6.68E-04 (± 4.15E-04)	4.18E-04 (± 2.22E-04)	4.78E-04 (± 2.18E-04)	5.98E-04 (± 3.39E-04)	2.06E-02
Bacillus	Min ; Max	***5.08E-05 ; 5.81E-02***	***1.20E-03 ; 5.62E-02***	4.30E-05 ; 8.90E-04	2.15E-03 ; 1.06E-02	2.31E-04 ; 9.35E-04	***3.57E-04 ; 5.90E-02***	1.46E-04 ; 1.19E-03	2.61E-04 ; 5.75E-04	3.25E-04 ; 6.32E-04	3.24E-04 ; 1.45E-03	2.06E-02
Jeotgalicoccus	Mean (± SD)	4.26E-03 (± 4.07E-03)	2.95E-03 (± 2.04E-03)	1.39E-04 (± 8.72E-05)	3.25E-03 (± 2.86E-04)	3.58E-05 (± 2.42E-05)	***9.65E-03 (± 1.21E-02)***	1.11E-02 (± 9.96E-03)	2.74E-03 (± 3.70E-04)	4.12E-03 (± 3.00E-04)	2.67E-03 (± 4.36E-04)	4.78E-03
Jeotgalicoccus	Min ; Max	3.81E-05 ; 2.14E-02	2.00E-04 ; 6.33E-03	7.73E-05 ; 2.01E-04	3.05E-03 ; 3.45E-03	0.00E+00 ; 7.43E-05	***2.28E-04 ; 8.10E-02***	8.99E-04 ; 3.57E-02	2.47E-03 ; 3.00E-03	3.91E-03 ; 4.33E-03	2.15E-03 ; 3.46E-03	4.78E-03
Lactobacillus	Mean (± SD)	***5.05E-03 (± 8.99E-03)***	2.02E-03 (± 3.15E-03)	4.88E-04 (± 4.59E-04)	7.37E-03 (± 2.62E-03)	1.67E-03 (± 8.29E-04)	5.41E-03 (± 7.65E-03)	1.47E-03 (± 1.11E-03)	9.15E-04 (± 4.20E-04)	4.16E-03 (± 8.80E-04)	1.35E-02 (± 4.16E-03)	3.80E-03
Lactobacillus	Min ; Max	***0.00E+00 ; 6.04E-02***	3.09E-04 ; 1.13E-02	1.63E-04 ; 8.13E-04	5.52E-03 ; 9.23E-03	8.54E-04 ; 3.16E-03	1.39E-04 ; 4.58E-02	1.39E-04 ; 4.15E-03	6.18E-04 ; 1.21E-03	3.54E-03 ; 4.78E-03	9.60E-03 ; 2.18E-02	3.80E-03
Paracocccus	Mean (± SD)	***7.48E-03 (± 1.20E-02)***	2.12E-03 (± 2.03E-03)	2.18E-04 (± 1.87E-04)	3.63E-04 (± 2.46E-04)	2.97E-05 (± 2.76E-05)	4.68E-03 (± 2.74E-03)	1.06E-03 (± 1.09E-03)	1.88E-04 (± 7.22E-05)	2.71E-04 (± 1.10E-04)	1.90E-04 (± 7.67E-05)	1.06E-03
Paracocccus	Min ; Max	***6.35E-05 ; 7.91E-02***	8.54E-05 ; 6.36E-03	8.59E-05 ; 3.50E-04	1.89E-04 ; 5.37E-04	0.00E+00 ; 8.63E-05	2.28E-04 ; 1.26E-02	0.00E+00 ; 2.66E-03	1.37E-04 ; 2.40E-04	1.93E-04 ; 3.48E-04	6.94E-05 ; 3.24E-04	1.06E-03
Rhodococcus	Mean (± SD)	6.25E-03 (± 6.89E-03)	3.36E-03 (± 2.63E-03)	3.04E-04 (± 8.06E-05)	2.74E-03 (± 1.13E-03)	2.89E-04 (± 1.74E-04)	***4.88E-03 (± 8.04E-03)***	2.11E-03 (± 2.72E-03)	6.45E-04 (± 1.40E-05)	1.54E-03 (± 1.05E-04)	2.73E-03 (± 2.58E-04)	1.80E-03
Rhodococcus	Min ; Max	2.54E-05 ; 3.33E-02	4.76E-04 ; 8.25E-03	2.47E-04 ; 3.61E-04	1.95E-03 ; 3.54E-03	1.16E-04 ; 5.89E-04	***2.95E-04 ; 6.53E-02***	2.46E-04 ; 9.25E-03	6.35E-04 ; 6.55E-04	1.46E-03 ; 1.61E-03	2.46E-03 ; 3.36E-03	1.80E-03
Moraxella	Mean (± SD)	***7.18E-03 (± 1.45E-02)***	4.32E-03 (± 4.99E-03)	3.17E-04 (± 8.32E-05)	2.04E-02 (± 2.17E-03)	1.74E-03 (± 1.13E-03)	2.85E-03 (± 4.24E-03)	2.07E-03 (± 3.66E-03)	1.09E-03 (± 4.99E-04)	4.65E-04 (± 3.71E-05)	7.52E-04 (± 2.25E-04)	3.92E-04

Moraxella	Min ; Max	***4.77E-05 ; 8.79E-02***	1.13E-04 ; 1.59E-02	2.58E-04 ; 3.75E-04	1.88E-02 ; 2.19E-02	5.20E-04 ; 3.96E-03	2.77E-05 ; 2.74E-02	0.00E+00 ; 1.02E-02	7.38E-04 ; 1.44E-03	4.39E-04 ; 4.91E-04	5.43E-04 ; 1.26E-03	3.92E-04
Macrococcus	Mean (± SD)	2.61E-03 (± 3.87E-03)	1.37E-03 (± 1.02E-03)	1.89E-04 (± 1.24E-05)	2.35E-03 (± 5.38E-05)	1.56E-04 (± 9.78E-05)	***7.32E-03 (± 1.05E-02)***	***8.01E-03 (± 2.35E-02)***	1.84E-03 (± 5.86E-04)	1.31E-03 (± 1.97E-04)	2.08E-03 (± 7.02E-04)	8.70E-04
Macrococcus	Min ; Max	1.23E-04 ; 2.46E-02	3.55E-04 ; 3.29E-03	1.80E-04 ; 1.98E-04	2.31E-03 ; 2.38E-03	4.42E-05 ; 3.30E-04	***8.04E-06 ; 7.08E-02***	***1.94E-04 ; 7.88E-02***	1.43E-03 ; 2.25E-03	1.17E-03 ; 1.45E-03	1.10E-03 ; 3.49E-03	8.70E-04
Rothia	Mean (± SD)	***6.49E-03 (± 1.64E-02)***	6.50E-03 (± 8.39E-03)	5.52E-04 (± 3.68E-04)	7.25E-03 (± 1.27E-03)	8.92E-04 (± 2.11E-04)	2.32E-03 (± 4.38E-03)	2.49E-03 (± 4.64E-03)	6.33E-04 (± 5.30E-04)	4.56E-04 (± 6.16E-05)	4.72E-04 (± 9.73E-05)	5.54E-04
Rothia	Min ; Max	***1.72E-05 ; 9.68E-02***	3.35E-04 ; 2.48E-02	2.92E-04 ; 8.13E-04	6.36E-03 ; 8.15E-03	5.68E-04 ; 1.35E-03	1.61E-05 ; 3.47E-02	8.83E-05 ; 1.50E-02	2.58E-04 ; 1.01E-03	4.13E-04 ; 5.00E-04	3.40E-04 ; 6.37E-04	5.54E-04
Bifidobacterium	Mean (± SD)	***4.38E-03 (± 1.01E-02)***	4.55E-03 (± 6.04E-03)	2.89E-04 (± 2.39E-04)	1.02E-02 (± 1.92E-03)	1.07E-03 (± 7.45E-04)	***3.54E-03 (± 1.20E-02)***	5.71E-03 (± 6.21E-03)	1.36E-03 (± 2.77E-04)	2.06E-03 (± 2.71E-04)	2.07E-03 (± 2.78E-04)	1.73E-03
Bifidobacterium	Min ; Max	***0.00E+00 ; 5.86E-02***	2.30E-04 ; 2.24E-02	1.20E-04 ; 4.58E-04	8.89E-03 ; 1.16E-02	3.97E-04 ; 2.55E-03	***0.00E+00 ; 8.57E-02***	1.23E-04 ; 1.86E-02	1.16E-03 ; 1.55E-03	1.87E-03 ; 2.25E-03	1.70E-03 ; 2.39E-03	1.73E-03
Flavobacterium	Mean (± SD)	4.46E-03 (± 7.18E-03)	2.29E-03 (± 2.55E-03)	5.61E-04 (± 6.11E-04)	2.45E-03 (± 1.08E-03)	1.69E-04 (± 9.03E-05)	***3.87E-03 (± 7.93E-03)***	5.08E-03 (± 7.39E-03)	2.44E-03 (± 6.24E-04)	7.92E-04 (± 1.40E-04)	2.35E-04 (± 9.40E-05)	7.51E-04
Flavobacterium	Min ; Max	0.00E+00 ; 3.79E-02	1.42E-04 ; 8.46E-03	1.29E-04 ; 9.93E-04	1.68E-03 ; 3.21E-03	7.16E-05 ; 3.61E-04	***8.56E-05 ; 6.66E-02***	1.72E-04 ; 2.53E-02	1.99E-03 ; 2.88E-03	6.93E-04 ; 8.90E-04	6.43E-05 ; 3.75E-04	7.51E-04
Brachybacterium	Mean (± SD)	***5.35E-03 (± 1.16E-02)***	2.78E-03 (± 3.73E-03)	9.78E-05 (± 6.54E-05)	1.55E-03 (± 2.35E-04)	1.69E-04 (± 9.42E-05)	3.39E-03 (± 4.81E-03)	1.53E-03 (± 1.24E-03)	6.04E-04 (± 1.02E-04)	5.87E-04 (± 1.37E-04)	1.06E-03 (± 2.74E-04)	1.82E-03
Brachybacterium	Min ; Max	***6.35E-05 ; 8.75E-02***	2.76E-04 ; 1.18E-02	5.15E-05 ; 1.44E-04	1.38E-03 ; 1.71E-03	3.98E-05 ; 3.16E-04	1.97E-04 ; 2.59E-02	9.24E-05 ; 3.45E-03	5.32E-04 ; 6.76E-04	4.90E-04 ; 6.84E-04	6.73E-04 ; 1.61E-03	1.82E-03
Enterococcus	Mean (± SD)	***4.19E-03 (± 9.85E-03)***	5.29E-03 (± 7.99E-03)	1.95E-04 (± 1.54E-04)	3.40E-03 (± 7.75E-04)	5.34E-04 (± 2.83E-04)	2.42E-03 (± 3.86E-03)	1.31E-03 (± 1.13E-03)	5.32E-04 (± 8.46E-05)	1.57E-03 (± 6.47E-05)	3.00E-03 (± 7.11E-04)	8.44E-04
Enterococcus	Min ; Max	***1.08E-04 ; 6.16E-02***	7.37E-04 ; 2.86E-02	8.59E-05 ; 3.03E-04	2.85E-03 ; 3.95E-03	2.74E-04 ; 1.22E-03	0.00E+00 ; 2.39E-02	6.16E-05 ; 3.84E-03	4.72E-04 ; 5.92E-04	1.52E-03 ; 1.61E-03	1.70E-03 ; 3.89E-03	8.44E-04

Pseudoclavibacter	Mean (± SD)	***5.73E-03 (± 1.24E-02)***	4.72E-03 (± 9.91E-03)	1.11E-04 (± 1.07E-05)	7.73E-04 (± 4.77E-04)	6.13E-05 (± 4.38E-05)	1.01E-03 (± 1.22E-03)	1.01E-03 (± 1.88E-03)	2.74E-04 (± 6.03E-05)	4.31E-04 (± 1.35E-04)	4.77E-04 (± 1.57E-04)	2.47E-04
Pseudoclavibacter	Min ; Max	***1.27E-05 ; 6.88E-02***	3.94E-05 ; 3.19E-02	1.03E-04 ; 1.18E-04	4.36E-04 ; 1.11E-03	7.36E-06 ; 1.29E-04	1.22E-04 ; 6.73E-03	2.86E-05 ; 6.56E-03	2.32E-04 ; 3.17E-04	3.35E-04 ; 5.26E-04	2.54E-04 ; 7.21E-04	2.47E-04
Clostridium sensu stricto 5	Mean (± SD)	1.45E-05 (± 4.79E-05)	1.14E-06 (± 2.63E-06)	0.00E+00 (± 0.00E+00)	3.00E-05 (± 1.92E-05)	1.64E-06 (± 4.91E-06)	***7.41E-03 (± 1.67E-02)***	2.79E-05 (± 3.37E-05)	0.00E+00 (± 0.00E+00)	0.00E+00 (± 0.00E+00)	7.11E-05 (± 1.23E-04)	1.11E-04
Clostridium sensu stricto 5	Min ; Max	0.00E+00 ; 3.47E-04	0.00E+00 ; 7.80E-06	0.00E+00 ; 0.00E+00	1.64E-05 ; 4.36E-05	0.00E+00 ; 1.47E-05	***0.00E+00 ; 9.54E-02***	0.00E+00 ; 1.07E-04	0.00E+00 ; 0.00E+00	0.00E+00 ; 0.00E+00	0.00E+00 ; 3.91E-04	1.11E-04
Clavibacter	Mean (± SD)	***5.45E-03 (± 1.08E-02)***	4.09E-03 (± 8.70E-03)	1.66E-04 (± 1.14E-04)	5.79E-04 (± 1.00E-04)	9.09E-05 (± 5.96E-05)	9.68E-04 (± 1.09E-03)	8.05E-04 (± 1.73E-03)	1.88E-04 (± 1.33E-04)	3.40E-04 (± 1.15E-04)	4.18E-04 (± 2.00E-04)	2.05E-04
Clavibacter	Min ; Max	***4.39E-05 ; 5.51E-02***	2.30E-04 ; 2.96E-02	8.59E-05 ; 2.47E-04	5.08E-04 ; 6.50E-04	3.18E-05 ; 2.12E-04	0.00E+00 ; 7.16E-03	4.01E-05 ; 5.97E-03	9.45E-05 ; 2.82E-04	2.58E-04 ; 4.21E-04	2.12E-04 ; 8.65E-04	2.05E-04
Carnobacterium	Mean (± SD)	***2.62E-03 (± 1.10E-02)***	2.41E-03 (± 3.78E-03)	3.74E-04 (± 2.24E-04)	2.23E-03 (± 1.14E-03)	2.37E-04 (± 8.65E-05)	3.00E-04 (± 6.73E-04)	***1.37E-02 (± 2.38E-02)***	8.65E-03 (± 1.88E-03)	1.18E-02 (± 2.86E-04)	4.57E-03 (± 1.22E-03)	1.14E-02
Carnobacterium	Min ; Max	***0.00E+00 ; 8.65E-02***	4.82E-05 ; 1.17E-02	2.16E-04 ; 5.33E-04	1.42E-03 ; 3.03E-03	1.27E-04 ; 3.61E-04	0.00E+00 ; 3.42E-03	***1.62E-04 ; 7.50E-02***	7.32E-03 ; 9.98E-03	1.16E-02 ; 1.20E-02	2.77E-03 ; 6.37E-03	1.14E-02
Atopococcus	Mean (± SD)	***2.11E-03 (± 6.92E-03)***	1.23E-03 (± 1.81E-03)	7.80E-05 (± 8.60E-05)	1.61E-03 (± 3.77E-04)	1.27E-04 (± 5.17E-05)	***3.35E-03 (± 2.01E-02)***	1.44E-03 (± 1.81E-03)	8.65E-04 (± 2.75E-05)	1.42E-03 (± 1.12E-04)	1.38E-03 (± 3.53E-04)	1.59E-03
Atopococcus	Min ; Max	***0.00E+00 ; 5.55E-02***	0.00E+00 ; 6.19E-03	1.72E-05 ; 1.39E-04	1.34E-03 ; 1.87E-03	4.05E-05 ; 2.12E-04	***0.00E+00 ; 1.65E-01***	1.44E-04 ; 5.32E-03	8.46E-04 ; 8.84E-04	1.34E-03 ; 1.50E-03	8.85E-04 ; 1.84E-03	1.59E-03
Aeromonas	Mean (± SD)	1.99E-03 (± 5.51E-03)	1.86E-03 (± 3.68E-03)	7.70E-04 (± 2.96E-04)	6.93E-04 (± 2.93E-04)	4.84E-04 (± 2.01E-04)	***1.76E-03 (± 7.84E-03)***	6.37E-04 (± 8.90E-04)	2.59E-04 (± 1.72E-04)	9.44E-04 (± 3.30E-05)	5.64E-03 (± 6.30E-04)	5.97E-04
Aeromonas	Min ; Max	0.00E+00 ; 4.30E-02	4.60E-05 ; 1.26E-02	5.61E-04 ; 9.79E-04	4.85E-04 ; 9.00E-04	2.41E-04 ; 8.11E-04	***0.00E+00 ; 6.28E-02***	0.00E+00 ; 2.49E-03	1.37E-04 ; 3.80E-04	9.21E-04 ; 9.68E-04	4.73E-03 ; 6.80E-03	5.97E-04
uncultured Fusobacterium sp.	Mean (± SD)	8.81E-04 (± 2.40E-03)	1.63E-03 (± 1.95E-03)	1.30E-04 (± 1.31E-05)	1.06E-02 (± 1.24E-03)	6.22E-04 (± 4.90E-04)	***2.61E-03 (± 1.14E-02)***	4.11E-03 (± 1.28E-02)	1.04E-03 (± 3.80E-04)	1.57E-03 (± 1.45E-04)	6.58E-04 (± 2.00E-04)	2.44E-03

uncultured Fusobacterium sp.	Min ; Max	0.00E+00 ; 1.56E-02	7.22E-05 ; 6.52E-03	1.20E-04 ; 1.39E-04	9.68E-03 ; 1.14E-02	1.57E-04 ; 1.66E-03	***0.00E+00 ; 6.97E-02***	0.00E+00 ; 4.27E-02	7.68E-04 ; 1.31E-03	1.47E-03 ; 1.68E-03	3.70E-04 ; 9.49E-04	2.44E-03
uncultured bacterium	Mean (± SD)	6.61E-04 (± 3.62E-03)	***1.17E-02 (± 3.27E-02)***	9.50E-04 (± 4.16E-05)	2.25E-02 (± 2.46E-03)	1.15E-03 (± 6.97E-04)	1.12E-03 (± 5.78E-03)	3.36E-04 (± 9.59E-04)	4.46E-05 (± 2.66E-05)	1.33E-03 (± 1.01E-04)	3.95E-04 (± 1.36E-04)	2.75E-03
uncultured bacterium	Min ; Max	0.00E+00 ; 2.89E-02	***0.00E+00 ; 1.09E-01***	9.21E-04 ; 9.79E-04	2.08E-02 ; 2.43E-02	3.26E-04 ; 2.49E-03	0.00E+00 ; 3.46E-02	0.00E+00 ; 3.21E-03	2.58E-05 ; 6.34E-05	1.26E-03 ; 1.41E-03	2.31E-04 ; 5.77E-04	2.75E-03
Anoxybacillus	Mean (± SD)	3.79E-04 (± 9.22E-04)	4.78E-05 (± 5.14E-05)	7.71E-06 (± 1.09E-05)	2.78E-03 (± 7.04E-04)	3.08E-03 (± 9.18E-04)	2.22E-04 (± 4.53E-04)	2.35E-05 (± 3.64E-05)	4.29E-06 (± 6.07E-06)	4.54E-05 (± 2.70E-05)	9.02E-05 (± 5.41E-05)	***2.36E-01***
Anoxybacillus	Min ; Max	0.00E+00 ; 5.62E-03	0.00E+00 ; 1.69E-04	0.00E+00 ; 1.54E-05	2.29E-03 ; 3.28E-03	1.73E-03 ; 4.51E-03	0.00E+00 ; 2.61E-03	0.00E+00 ; 1.06E-04	0.00E+00 ; 8.59E-06	2.63E-05 ; 6.45E-05	1.79E-05 ; 1.94E-04	***2.36E-01***
Empedobacter	Mean (± SD)	***2.16E-03 (± 7.65E-03)***	4.28E-04 (± 9.11E-04)	2.96E-03 (± 4.12E-03)	6.44E-04 (± 8.94E-05)	2.08E-06 (± 4.28E-06)	1.21E-03 (± 3.07E-03)	3.58E-03 (± 1.04E-02)	7.51E-03 (± 2.01E-03)	2.55E-03 (± 7.39E-04)	1.31E-04 (± 7.37E-05)	4.95E-04
Empedobacter	Min ; Max	***0.00E+00 ; 6.08E-02***	0.00E+00 ; 3.09E-03	4.30E-05 ; 5.87E-03	5.81E-04 ; 7.07E-04	0.00E+00 ; 1.16E-05	0.00E+00 ; 2.15E-02	0.00E+00 ; 3.50E-02	6.09E-03 ; 8.93E-03	2.03E-03 ; 3.07E-03	4.59E-05 ; 2.69E-04	4.95E-04
Stenotrophomonas	Mean (± SD)	1.38E-03 (± 2.87E-03)	1.37E-03 (± 1.44E-03)	3.06E-04 (± 3.60E-04)	6.33E-04 (± 5.22E-04)	6.74E-05 (± 6.59E-05)	***2.65E-03 (± 7.96E-03)***	6.36E-04 (± 8.88E-04)	1.66E-04 (± 1.74E-04)	7.15E-05 (± 2.67E-05)	6.68E-05 (± 5.87E-05)	9.38E-05
Stenotrophomonas	Min ; Max	0.00E+00 ; 1.90E-02	6.57E-05 ; 4.18E-03	5.15E-05 ; 5.61E-04	2.63E-04 ; 1.00E-03	0.00E+00 ; 2.04E-04	***0.00E+00 ; 6.48E-02***	0.00E+00 ; 2.81E-03	4.29E-05 ; 2.89E-04	5.26E-05 ; 9.03E-05	0.00E+00 ; 1.35E-04	9.38E-05
Exiguobacterium	Mean (± SD)	***1.80E-03 (± 1.07E-02)***	2.59E-03 (± 7.02E-03)	1.65E-02 (± 1.90E-03)	1.65E-02 (± 2.33E-03)	4.82E-04 (± 1.84E-04)	1.86E-04 (± 3.78E-04)	5.67E-04 (± 1.84E-03)	5.39E-05 (± 3.43E-06)	3.69E-05 (± 2.56E-06)	1.52E-05 (± 2.01E-05)	1.71E-05
Exiguobacterium	Min ; Max	***0.00E+00 ; 8.63E-02***	0.00E+00 ; 2.33E-02	1.52E-02 ; 1.79E-02	1.49E-02 ; 1.82E-02	2.58E-04 ; 7.85E-04	0.00E+00 ; 2.59E-03	0.00E+00 ; 6.12E-03	5.15E-05 ; 5.64E-05	3.51E-05 ; 3.87E-05	0.00E+00 ; 5.83E-05	1.71E-05
Porphyromonas	Mean (± SD)	***1.40E-03 (± 7.46E-03)***	1.17E-03 (± 1.36E-03)	1.36E-04 (± 2.64E-06)	4.22E-03 (± 4.14E-04)	2.46E-04 (± 1.44E-04)	1.65E-03 (± 5.66E-03)	4.42E-04 (± 5.40E-04)	6.53E-05 (± 7.32E-06)	1.94E-04 (± 2.00E-04)	7.49E-05 (± 3.32E-05)	1.88E-04
Porphyromonas	Min ; Max	***0.00E+00 ; 6.13E-02***	1.86E-04 ; 4.65E-03	1.34E-04 ; 1.37E-04	3.93E-03 ; 4.52E-03	8.75E-05 ; 4.61E-04	0.00E+00 ; 4.39E-02	1.15E-05 ; 1.48E-03	6.01E-05 ; 7.05E-05	5.26E-05 ; 3.35E-04	2.50E-05 ; 1.17E-04	1.88E-04

Yersinia	Mean (± SD)	***1.52E-03 (± 7.13E-03)***	5.59E-04 (± 8.49E-04)	2.03E-03 (± 2.66E-03)	6.02E-04 (± 5.03E-04)	4.43E-04 (± 2.77E-04)	2.93E-04 (± 1.63E-03)	2.29E-03 (± 4.51E-03)	1.46E-03 (± 8.69E-04)	1.05E-03 (± 3.36E-04)	2.92E-03 (± 4.42E-04)	7.16E-04
Yersinia	Min ; Max	***0.00E+00 ; 5.63E-02***	0.00E+00 ; 2.82E-03	1.49E-04 ; 3.91E-03	2.47E-04 ; 9.58E-04	1.62E-04 ; 1.04E-03	0.00E+00 ; 1.31E-02	1.38E-05 ; 1.49E-02	8.50E-04 ; 2.08E-03	8.16E-04 ; 1.29E-03	2.31E-03 ; 3.75E-03	7.16E-04
Brochothrix	Mean (± SD)	1.79E-05 (± 1.06E-04)	4.95E-05 (± 9.22E-05)	2.92E-05 (± 1.95E-05)	0.00E+00 (± 0.00E+00)	0.00E+00 (± 0.00E+00)	1.75E-06 (± 6.23E-06)	***1.49E-02 (± 4.71E-02)***	1.38E-03 (± 4.54E-04)	1.57E-03 (± 1.35E-04)	5.50E-04 (± 2.89E-04)	2.56E-04
Brochothrix	Min ; Max	0.00E+00 ; 8.51E-04	0.00E+00 ; 2.73E-04	1.54E-05 ; 4.30E-05	0.00E+00 ; 0.00E+00	0.00E+00 ; 0.00E+00	0.00E+00 ; 2.98E-05	***0.00E+00 ; 1.57E-01***	1.06E-03 ; 1.70E-03	1.47E-03 ; 1.66E-03	1.65E-04 ; 9.74E-04	2.56E-04
Hafnia	Mean (± SD)	***1.91E-03 (± 1.25E-02)***	1.79E-03 (± 5.43E-03)	9.82E-04 (± 3.49E-04)	8.31E-04 (± 2.21E-04)	2.30E-04 (± 1.26E-04)	1.33E-04 (± 6.61E-04)	8.86E-05 (± 1.77E-04)	4.92E-05 (± 3.97E-05)	9.96E-05 (± 4.16E-05)	3.67E-04 (± 9.90E-05)	0.00E+00
Hafnia	Min ; Max	***0.00E+00 ; 1.01E-01***	0.00E+00 ; 1.81E-02	7.35E-04 ; 1.23E-03	6.74E-04 ; 9.87E-04	8.10E-05 ; 4.32E-04	0.00E+00 ; 3.91E-03	0.00E+00 ; 5.90E-04	2.11E-05 ; 7.73E-05	7.02E-05 ; 1.29E-04	2.12E-04 ; 5.36E-04	0.00E+00
Others	Mean (± SD)	***4.04E-01 (± 1.57E-01)***	***3.53E-01 (± 1.41E-01)***	3.05E-02 (± 1.76E-02)	***2.65E-01 (± 8.25E-04)***	1.60E-02 (± 6.73E-03)	***5.24E-01 (± 1.47E-01)***	***2.03E-01 (± 1.32E-01)***	***5.41E-02 (± 9.83E-04)***	***1.12E-01 (± 2.27E-03)***	***1.13E-01 (± 1.61E-02)***	***1.80E-01***
Others	Min ; Max	***9.96E-03 ; 6.90E-01***	***6.35E-02 ; 5.14E-01***	1.81E-02 ; 4.29E-02	***2.65E-01 ; 2.66E-01***	9.08E-03 ; 2.98E-02	***3.87E-02 ; 7.49E-01***	***2.78E-02 ; 4.11E-01***	***5.34E-02 ; 5.48E-02***	***1.11E-01 ; 1.14E-01***	***9.33E-02 ; 1.32E-01***	***1.80E-01***

Table shows 47 genera > 5% relative abundance in at least one sample, and Others (genera < 5% relative abundance), their mean (± SD) relative abundance per group, as well as their minimum and maximum relative abundance per group. Bold italics show group(s) in which the genera were present in greater than 5% relative abundance in at least one sample. Genera are listed in order of cumulative relative abundance for all samples. For each group, bulk tanks (BT) n = 67, collection tankers (CT) n = 11, whole milk silos (WMS) n = 2, skimmed milk silos (SMS) n = 2, skimmed milk powder (SMP) n = 9, cream n = 1.

Chapter 5. Environmental monitoring of a dairy processing facility using multiple high throughput sequencing methods

Authors: Aoife J. McHugh, Min Yap, Fiona Crispie, Conor Feehily, Colin Hill and Paul D. Cotter

Contributions: Candidate performed swabbing, culture-based analysis, DNA extraction, DNA library preparations, MinION sequencing, bioinformatic and statistical analysis.

Candidate drafted and edited chapter.

MY helped candidate with swabbing and culture-based lab work.

FC helped candidate with MinION sequencing.

CF, CH and PDC supervised the candidate.

5.1 Abstract

Food processing environments can harbor microorganisms responsible for food contamination and spoilage, potentially leading to foodborne disease and/or economic losses. Efficient and accurate identification of undesirable microorganisms throughout the food chain, including food production, food processing and the resultant foods, can allow the identification and elimination of sources of contamination and the implementation of control measures. Currently, microbial monitoring of the food chain relies heavily on culture-based techniques. These assays are selected on the basis of the microbes expected to be present in the environment, and thus do not cater for unexpected contaminants. It is also notable that many culture-based assays are also unable to distinguish between undesirable taxa and closely related harmless strains. Furthermore, even when multiple culture-based approaches are used in parallel, it is still not possible a comprehensively characterize the entire microbiology of a food-chain sample.

High throughput DNA sequencing represents a potential means through which microbial monitoring of the food chain can be enhanced. While the corresponding sequencing platforms, such as the Illumina MiSeq, NextSeq and NovaSeq, are most typically found in research or commercial sequencing laboratories, newer portable platforms, such as the Oxford Nanopore MinION, offer the potential for rapid analysis of food chain microbiomes. In this study, having initially assessed the ability of MinION-based sequencing to discriminate between different components of a mock metagenomic mixture of related food spoilage spore-forming microorganisms, we proceeded to compare outcomes from Oxford Nanopore and

Illumina sequencing of metagenomic DNA obtained from environmental monitoring in an active food processing facility.

Overall, Oxford Nanopore MinION sequencing provided accurate classification to species level, which was comparable to Illumina-derived outputs. However, while the MinION-based approach provided a means of easy library preparations and portability, the high concentrations of DNA needed to run the rapid sequencing protocols can be a limiting factor.

5.2 Introduction

Dairy processing environments harbor microorganisms that have the potential to contaminate food before and during processing (Gleeson, O'Connell and Jordan, 2013; Doyle *et al.*, 2017; Faille *et al.*, 2014; Wang *et al.*, 2019; Fysun *et al.*, 2019). Some of these microorganisms have the potential to cause spoilage or be pathogenic (Doyle *et al.*, 2015; Cho *et al.*, 2018; Sadiq *et al.*, 2016; Burgess, Lindsay and Flint, 2010). Routine environmental monitoring is carried out in food processing environments for this reason, and usually involves the use of swabbing and agar plating to determine total numbers of general (e.g., total bacteria count) or specific (generally potentially spoilage-associated or pathogenic species) categories of microorganisms (Cho *et al.*, 2018). These analyses frequently involve phenotype-based agar assays, some of which can yield high false positive rates (Doyle, O'Toole and Cotter, 2018; Tallent *et al.*, 2012). These approaches are limited by the fact that they do not provide information about non-target species or, indeed, the microbial population as a whole.

DNA sequencing methods have recently been applied to dairy and environmental samples to determine microbial population composition and allow source tracking (Doyle *et al.*, 2017; McHugh *et al.*, 2018; Fretin *et al.*, 2018; Cho *et al.*, 2018). High throughput metagenomic sequencing can provide greater insights into the taxonomic composition of populations present in these environments than culture based methods and can also provide information relating to the functional, including virulence and spoilage, potential of species and strains present. Despite these benefits, high throughput metagenomic sequencing approaches typically

require expensive reagents and platforms as well as personnel skilled in molecular biology and bioinformatics skills for data generation and interpretation that limit their routine implementation in manufacturing facilities. Some of these issues have the potential to be addressed through use of portable DNA sequencing devices such as the Oxford Nanopore MinION. The MinION's long read sequences have the potential to provide more streamlined, user-friendly analysis options, which could allow easier detection and identification of causative agents of contamination. Such approaches have been previously used in a clinical setting to identify causative agents of disease from metagenomic samples (Charalampous *et al.*, 2019), including some studies in which the results generated were compared with those generated through Illumina sequencing (Quick *et al.*, 2017; Kafetzopoulou *et al.*, 2018) or culture-based analysis (Sanderson *et al.*, 2018), but have yet to be applied to food processing settings for environmental monitoring.

To this end, we carried out an initial proof-of-concept study to determine the ability of Oxford Nanopore MinION rapid sequencing to correctly classify a simple, four strain, mock community of highly related spore-forming microorganisms of relevance to the dairy processing chain. Prompted by this initial analysis, we proceeded to compare the ability of Oxford Nanopore rapid sequencing, Illumina-based and culture-based methods to characterize the microbiota of environmental swabs collected from a food processing facility. Overall, MinION sequencing performed well, with very few differences in species level taxonomic classification when compared to Illumina sequencing of equivalent samples. However, limitations relating to sourcing DNA of sufficient high quantity and high quality for routine

MinION sequencing were noted. Regardless, the potential benefits of the routine application of metagenomic sequencing to food processing environments were clear.

5.3 Materials and methods

5.3.1 Mock community

DNA from 4 target strains, *Bacillus cereus* DSM 31/ATCC 14579, *Bacillus thuringiensis* DSM 2046/ATCC 10792, *Bacillus licheniformis* DSM 13/ATCC 14580 and *Geobacillus stearothermophilus* DSM 458 (Accession numbers GCF_000007825.1, GCF_002119445.1, GCF_000011645.1 and GCF_002300135.1, respectively), was combined to represent a 'mock' metagenomic sample of spore-forming bacteria. Genomic DNA was purchased (*B*. *licheniformis* and *G*. *stearothermophilus*; DSMZ) or extracted from in-house stocks (*B*. *cereus* and *B*. *thuringiensis*). Where necessary, DNA extraction was performed using the GenElute Bacterial Genomic DNA extraction kit (Sigma Aldrich, NA2110) according to manufacturer's instructions for Gram positive bacteria DNA extraction except that DNA was eluted in 75 µl elution solution with a view to increasing the concentration of genomic DNA generated. DNA was stored at -20°C after which DNA concentrations were determined using the Qubit double-stranded DNA high sensitivity (HS) assay kit (BioSciences) and ran on 1% agarose gel to check quality. DNA was diluted to 24 pM and pooled at equimolar concentrations before storing at -20°C. 16S rRNA metagenome sequencing, using the 16S rapid barcoding kit SQK-RAB204, as well as rapid whole metagenome sequencing (WMGS), using the rapid sequencing kit SQK-RAD004, was performed using the Oxford Nanopore MinION sequencer. These kits required 10 ng and 400 ng of DNA input, respectively. More specifically, for 16S rRNA MinION sequencing, the SQK-RAB204 16S rapid barcoding kit was used for library preparation according to manufacturer's instructions with

barcode 01. DNA was sequenced on FloMIN 106 R9 version flowcell mk1 with minKNOW version 1.7.14 according to manufacturer's instructions. For rapid whole metagenome MinION sequencing, the SQK-RAD004 rapid sequencing kit was used to prepare the DNA according to manufacturer's instructions, DNA was sequenced on FloMIN 106 R9 version flowcell mk1 with MinKNOW version 1.11.5 according to manufacturer's instructions.

5.3.1.1 Bioinformatic analysis of mock community metagenomic DNA

Genome sequences for the 4 strains represented in the mock metagenomic community were downloaded from NCBI RefSeq and aligned in a pairwise manner using the Artemis comparison tool (ACT) (Carver *et al.*, 2005) (Supplemental Figure 5.1). 16S DNA sequences were rebasecalled using Albacore version 2.2.6. FastQC was used to check sequence length and quality. IDBA fq2fa was used to convert fastq files to fasta format. BLASTn alignment (Altschul *et al.*, 1990) of sequences against 16S Silva database (release 132) (Pruesse *et al.*, 2007; Quast *et al.*, 2012) was performed with taxonomic classification by MEGAN (version 6.12.3) blast2rma followed by rma2info with rank included (Huson *et al.*, 2007). Genus and species levels of classification were determined, and relative abundances calculated and plotted using R ggplot2 (Wickham, 2009). Following basecalling with Albacore, Porechop version 0.2.4 was used to remove adaptors from rapid WMGS reads before FastQC was used to check sequence length and quality and IDBA fq2fa was used to convert fastq format to fasta format (Peng *et al.*, 2012). LAST alignment of reads (Kielbasa *et al.*, 2011; Sheetlin *et al.*, 2014) was performed against the NR database (March 2018) (Pruitt, Tatusova and Maglott, 2005; Pruitt *et al.*, 2012) with

MEGAN long read (LR) (MEGAN version 6.12.3) (Huson *et al.*, 2018) taxonomic classification applied using daa meganizer and daa2info with rank information included. Ranks were split, relative abundances calculated and plotted using R ggplot2 (Wickham, 2009).

The assembly of contigs from metagenomic reads was performed using Canu version 1.7 (Koren *et al.*, 2017) with -nanopore-raw flag. MUMmer alignment was performed on the assembled contigs against the 4 known species genomes from RefSeq, with dnadiff used to highlight differences between assemblies and reference genomes (Kurtz *et al.*, 2004; Delcher *et al.*, 2002). The resulting comparisons were visualised using R ggbio and GenomicRanges (Yin, Cook and Lawrence, 2012; Lawrence *et al.*, 2013).

5.3.2 Environmental sample collection and processing

Environmental swabbing was performed in a commercial dairy processing pilot plant. Eight locations were swabbed during the course of a single day, after cleaning in place (CIP) had been completed and before the next round of dairy processing (Figure 5.1). These eight locations included a table, door, wall, gaskets/flow plate seals, external surface of dryer balance tank, internal surface of dryer balance tank, external surface of evaporator, and drain beside evaporator. Overall, these eight locations were swabbed over three different months (October, November, December), at a frequency of once per month.

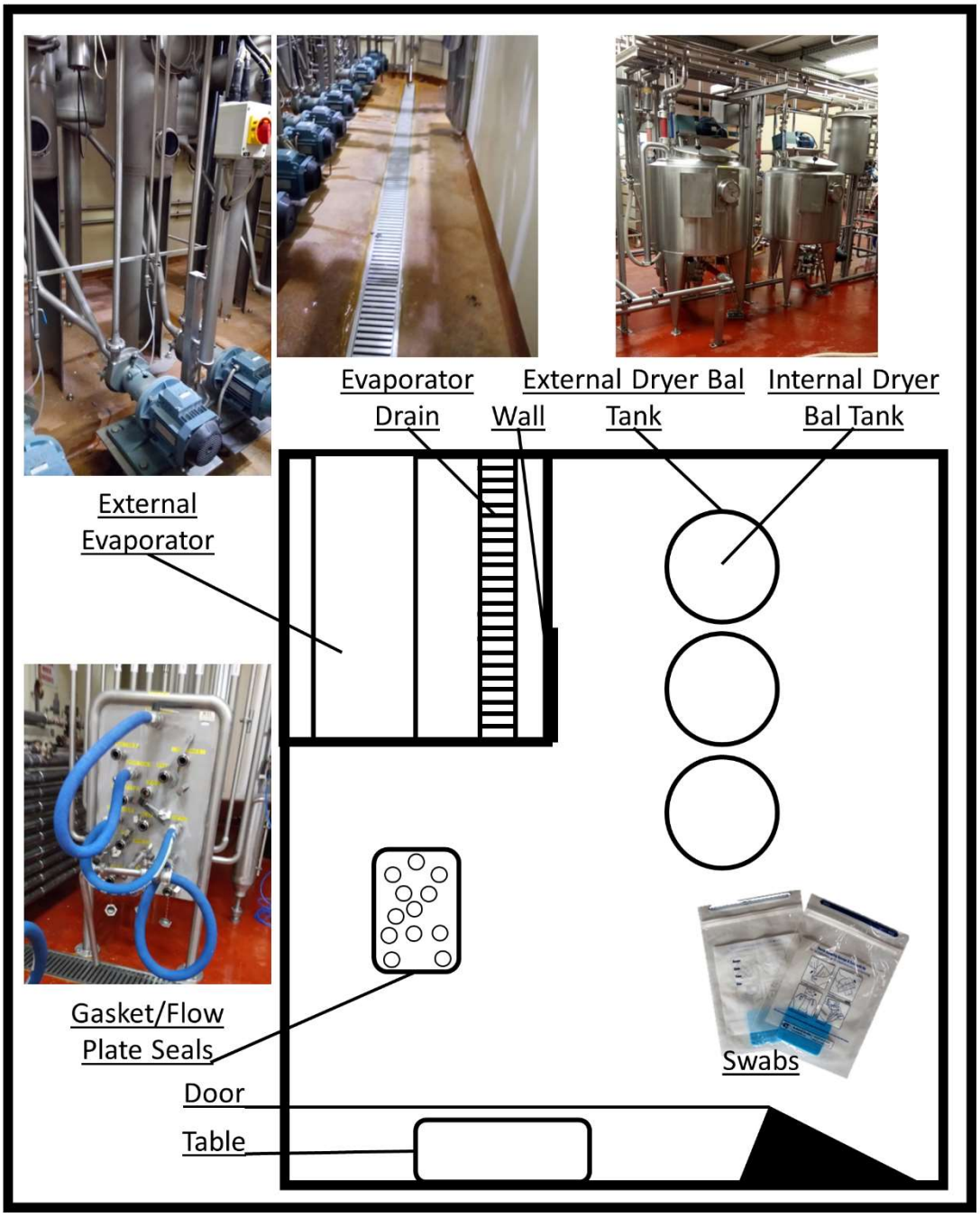


Figure 5.1. Schematic of dairy processing facility sampling areas.

Dairy processing facility schematic includes the 8 areas sampled in each of October, November, and December 2018. Areas were sampled post CIP and prior to the recommencement of processing.

Swabbing was performed using Technical Service Consultants Ltd. sponges in neutralising buffer (Sparks Lab Supplies, SWA2023), and swabbing was performed according to manufacturer's instructions (hygiene sponge sampling kits swabbing procedure). Briefly, a stomacher bag containing a pre-moistened sponge was shaken to bring the sponge to the bottom of the bag. The bag was torn open above the zip lock; then, holding the sponge at the bottom from the outside, the bag was carefully peeled back, from above the zip lock over the gloved hand, taking care not to touch the inside of the bag or sponge. The exposed sponge was the used to swab an area of 360 cm^2 swabbing vertically with one side of the sponge and horizontally over the same area with the other side of the sponge. The bag was then carefully reverted to its original position, without touching the inside and the zip lock sealed. 5 swabs of each area to be sampled were performed in this way. The surface area was then wiped with disinfectant to remove neutralising buffer. In the laboratory, the 5 sponges for each area were pooled aseptically into the stomacher bag of one of the sponges. Each bag of 5 sponges was subjected to stomaching at 260 rpm for 1 minute. The liquid was then removed, yielding 21 ml for each sample of 5 sponges. 20 ml was prepared for DNA extraction. 1 ml was used for culturing. In order to prepare sample for DNA extraction, 2 x 15 ml falcon tubes for each sample holding a total of 20 ml were centrifuged at 4,500 x g for 20 min at 4°C. The supernatant was discarded, and pellet resuspended in 500 µl UV treated, autoclaved phosphate buffered saline (PBS). The two resuspended pellets for each area were pooled into a 2 ml microfuge tube. This tube was centrifuged at 13,000 x g for 2 min and the supernatant was discarded. The pellet was stored at -80°C for up to 1 month before DNA extraction. Swab negative controls were also processed

in the same way for each sampling day. Briefly, 5 swabs were pooled, subjected to stomaching, liquid collected, 1 ml split for culturing, 20 ml pelleted, washed and frozen.

5.3.2.1 Culture analysis

Of the 1 ml of liquid recovered from each stomacher bag, 100 µl was plated on BHI agar in triplicate. Another 100 µl was used for serial tenfold dilution and spread plate on BHI agar in triplicate. All agar plates incubated at 30°C for 48 h. 600 µl of liquid was subjected to spore pasteurisation by heating to 80°C for 12 min in a heating block. This heat treated liquid was then spread plated on BHI in triplicate for incubation at both 30°C and 55°C for 48 h (6 plates per sample), after which time colonies were counted to determine colony forming units CFU ml^{-1} and the agar plates stored at 4°C for further use.

For each sample, the colonies from one agar plate, onto which the neat stomacher bag liquid had been plated, were removed by washing and pelleted to facilitate DNA extraction to represent metagenomic DNA from easy to culture environmental microorganisms. To this end, 5 ml PBS was added to the agar plate, and swirled around, before colonies were scraped off with a sterile Lazy-L spreader (Sigma-Aldrich) and 4 ml recovered into a sterile 15 ml falcon tube. This was centrifuged at 4,500 x g for 20 min at 4°C before removing supernatant. The resulting pellet was resuspended in 1 ml PBS and transferred to a 2 ml microfuge tube. The tube was centrifuged at 13,000 x g for 2 min at room temperature and supernatant removed. The pellet was stored at -80°C for up to three months before DNA extraction. From other agar plates, isolated colonies with obviously different morphologies from

each sample were picked, restreaked for purity, inoculated in BHI broth and stocked at -20°C in a final concentration of 25% glycerol, for future use.

5.3.2.2 DNA extraction and MDA amplification

The Qiagen PowerSoil Pro kit was used for DNA extractions from both environmental sample pellets, and easily culturable washed plate pellets. Easily culturable pellets were removed from -80°C storage and resuspended in 1 ml PBS. 200 µl (or 500 µl for 9 smaller pellets, corresponding door, external evaporator and internal dryer balance tank samples for all 3 months) was removed and centrifuged at 12,000 x g for 2 min. The supernatant was discarded and the pellet retained. These pellets, and those sourced directly from environmental swabbing, i.e. without culture, were resuspended in 800 µl CD1 and transferred to a Powerbead Pro tube, which had been centrifuged briefly to ensure beads were at the bottom. Powerbead Pro tubes were secured in a tissue lyser set at 20 Hz for 10 min before centrifuging and following the rest of the PowerSoil Pro kit manufacturer's instructions, eluting in a smaller volume, of 35 µl, to maximise concentration. Elution solution was allowed to sit on the membrane for 5 min before centrifugation and DNA collection. DNA was stored at -20°C. For each sampling day, negative controls, involving unused swabs, were also prepared by following an identical extraction protocol and additional negative controls, to detect kit contaminants, were generated whereby an extraction was performed using the kit reagents alone, starting with 800 µL solution CD1.

Whole metagenome amplification was performed using multiple displacement amplification (MDA) with the REPLI-g Single Cell kit (Qiagen, 150345). MDA was

performed using DNA from environmental samples and controls for each day. These controls consisted of swab negative control, DNA extraction kit negative control, blank MDA preparation as a MDA negative control and mock metagenomic community (section 5.3.1) as a positive control. DNA concentrations were determined using Qubit HS kit. Samples with high DNA concentrations were diluted such that all samples had a final concentration of < 10 ng in 2.5 µl. MDA amplification was performed according to manufacturer's instructions. Briefly, 500 µl of H_2O sc was added to buffer DLB, mixed well and centrifuged briefly, storing at -20°C for up to 6 months. Buffer D1 and N1 were prepared according to instructions for 12 sample amplifications at a time (8 environmental samples, 1 positive control, 3 negative controls (swab, extraction, MDA), and prepared fresh on the day of use. 2.5 µl buffer D1 was added to 2.5 µl DNA, this was mixed by vortexing and centrifuged briefly. Samples were incubated at room temperature for 3 min. 5 µl buffer N1 was added, mixed by vortexing and centrifuged briefly before storing on ice. Mastermix was prepared on ice according to manufacturer's instructions. For each amplification 40 µl master mix was added to 10 µl denatured DNA. This was incubated on thermocycler at 30°C for 8 h. The polymerase was then inactivated by heating samples to 65°C for 3 min also in the thermocycler. Thermocycler settings (2 x 4 h holds at 30°C and 1 x 3 min hold at 65°C.) Amplified DNA was then stored at -20°C.

5.3.2.3 Library preparation and sequencing

5.3.2.3.1 MinION library

DNA concentrations of 36 MDA samples were measured using both the Qubit broad range (BR) and HS assays and diluted to 400 ng in 7.5 µl. Three libraries were prepared, containing 12 samples each (8 environmental MDA samples, 3 MDA negative controls (swab, extraction and MDA kit negative controls) and a MDA mock community positive control) per flow cell. The SQK-RBK004 rapid barcoding kit was used to prepare the DNA according to manufacturer's instructions, including an optional Ampure XP clean up step, directly prior to sequencing. DNA was sequenced on FloMIN 106 R9 version flowcell mk1 with MinKNOW version 18.12.4 according to manufacturer's instructions.

5.3.2.3.2 Illumina Nextera library

The DNA concentrations of MDA (36), non-MDA (i.e., metagenomic DNA not subjected to pre-processing (NPP)) (33), and easily culturable (Plate) (24) metagenomic DNA samples was measured using the Qubit HS kit and diluted. DNA was prepared for Illumina sequencing following Illumina Nextera XT Library Preparation Kit guidelines except that tagmentation was performed for 7 min. DNA tagmentation was visualised using Agilent, and average fragment size calculated. The DNA concentration was measured by Qubit HS and the concentration then calculated, before diluting and pooling at equimolar ratios. The DNA library was sequenced on Illumina NextSeq at the Teagasc DNA sequencing facility, with a NextSeq (500/500) High Output 300 cycles v2.5 kit (Illumina 20024908).

5.3.2.3.3 16S rDNA Sanger sequencing of isolated colonies

16S colony PCR was performed using universal primers 27F and 338R for 16S gene (AGAGTTTGATCCTGGCTCAG and CATGCTGCCTCCCGTAGGAGT, respectively). Colonies were picked, and mixed in 50 µl PCR water, before microwaving on full power for 1 minute to disrupt cells. Master mix consisting of 5 µl AccuTaq LA 10x buffer, 2.5 µl 10 mM dNTP mix, 1 µl DMSO, 2 µl 10 µM Forward primer, 2 µl 10 µM reverse primer, 32 µl PCR water, 0.5 µl AccuTaq LA DNA polymerase (Sigma Aldrich, D8045) per amplification was made. 45 µl mastermix was added to each tube of 5 µl disrupted colony, before centrifuging briefly to mix and placing on pre-programmed thermocycler with 95°C x 5 min, 25 cycles of 95°C x 30 sec, 55°C x 30 sec, 72°C x 30 sec and a final 72°C x 5 min, before holding at 4°C. PCR products were run on a 1 % agarose gel, before cleaning with 1.8 x Ampure XP. 5 µl of each cleaned up PCR product was aliquoted into a 96 well plate and 5 µl of forward primer added on top at 5 µM according to GATC requirements. A unique barcode was added to each plate and sent to GATC Biotech for Sanger sequencing. A subset of amplicons were also sequenced with the reverse primer to ensure accuracy.

5.3.2.4 Bioinformatic analysis of environmental metagenomic DNA

5.3.2.4.1 Analysis of MinION data

Guppy basecalled reads obtained from MinKnow (version 18.12.4) were demultiplexed using Guppy barcoder version (2.1.3) to produce a barcoding summary text file. This contained the percentage match of each read to which a barcode had been assigned (with a minimum score of 60, the default). All fastq files produced by MinKnow were concatenated and guppy_bcsplit.py obtained from

website 1 allowed demultiplexing of reads based on their barcode assigned in the barcoding summary text file. Porechop (version 0.2.4) was used to remove adaptors from rapid kit sequence reads before Fastqc was used to check sequence length and quality and Idba fq2fa was used to convert fastq to fasta (Peng *et al.*, 2012). LAST alignment of fasta files (Kielbasa *et al.*, 2011; Sheetlin *et al.*, 2014) against the nr database (March 2018) (Pruitt, Tatusova and Maglott, 2005; Pruitt *et al.*, 2012) was performed with the MEGAN LR classification (MEGAN version 6.12.3) (Huson *et al.*, 2018) using daa meganizer and daa2info with rank. Files were merged, ranks were split, total number of bases sequenced, and classified were calculated. Relative abundances calculated and plotted using R ggplot2 (Wickham, 2009).

5.3.2.4.2 Analysis of NextSeq data

BCl2fastq was used to convert raw sequence reads from Illumina NextSeq to fastq format. Kneaddata from bioBakery (McIver *et al.*, 2018) used trimmomatic for quality filtering and trimming paired end files (Bolger, Lohse and Usadel, 2014) with BMTagger to remove human and bovine reads. FastQC was used to visualise sequence length and quality. Idba converted fastq to fasta (Peng *et al.*, 2012). Diamond alignment (Buchfink, Xie and Huson, 2015) of fasta files was performed against the nr database (march 2018) (Pruitt, Tatusova and Maglott, 2005; Pruitt *et al.*, 2012) with MEGAN classification (MEGAN version 6.12.3) (Huson *et al.*, 2018) performed using daa meganizer and daa2info with rank information included. Files were merged, ranks were split, total number of bases sequenced and classified calculated and relative abundances calculated and plotted using R ggplot2 (Wickham, 2009).

Illumina data was also analysed using Kraken2 and Bracken (Lu *et al.*, 2017; Wood and Salzberg, 2014) for taxonomy classification as well as using metaphlan2 (Truong *et al.*, 2015) for taxonomy classification for the purpose of comparison.

5.3.2.4.2.1 Generation of MinION-Illumina hybrid Metagenome-assembled genomes

Metagenome assembled genomes (MAGs) were generated as follows. MDA amplified sequences from both Illumina and Oxford Nanopore sequencing were assembled using OPERA-MS (Bertrand *et al.*, 2019). Illumina reads were then mapped against assemblies using bowtie2 (Langmead and Salzberg, 2012) and bam files sorted using samtools (Li *et al.*, 2009). Depth was calculated and Metabat2 ran on assembled contigs to produce bins (Kang *et al.*, 2015; Kang *et al.*, 2019). Checkm was used to determine the quality of the MAGs. Prokka (Seemann, 2014) was used to generate .ffn files from bins, Kaiju (Menzel, Ng and Krogh, 2016)-based taxonomic classification was performed on the open reading frames from prokka. Megan LR (Huson *et al.*, 2018) was also used on the whole bins for taxonomic classification of high quality MAGs.

5.3.2.4.3 Culture- and 16S rRNA Sanger sequence-based analysis

CFUs were determined on the basis of an average of three agar plates per sample. CFU per swab was calculated by dividing by 5 (5 swabs=1 sample, and each swab covered area 360cm^2). 16S rRNA Sanger sequences resulting from morphologically different isolates per sample were blasted using BLASTn against the 16S ribosomal RNA (Bacteria and Archaea) database on NCBI, with top hits recorded, and genus level classification analysed.

5.3.2.4.4 Statistics

Pairwise Wilcoxon rank sums test using Benjamini Hochberg *p*-value correction analysis was used to compare sample groups, including investigations of the impact of sequencer type on taxonomy classification with MinION MDA-treated and NextSeq MDA-treated samples. The impact of MDA amplification was also investigated in this way through comparison between NextSeq MDA treated samples and NextSeq no pre-processing (NPP) samples. Differences in taxonomy classification between sequences derived from environmental metagenomic DNA versus those sourced from easy to culture microorganisms was shown by comparing NextSeq NPP and NextSeq easy to culture (plate) sequences.

5.4 Results

5.4.1 MinION sequencing accurately identifies, and distinguishes between, genomic DNA from four related, dairy environment-associated, sporeformers

Metagenomic DNA representing a simple mock community of 4 related dairy processing-associated, spore-forming contaminants, i.e., *B*. *cereus*, *B*. *thuringiensis*, *B*. *licheniformis*, *G*. *stearothermophilus*, was sequenced using Oxford Nanopore MinION rapid sequencing kits. This proof-of-concept exercise was performed to determine the extent to which MinION-based sequencing could identify, and discriminate between, related, and in some cases difficult to distinguish, microorganisms found in dairy processing environments. 16S rRNA-based sequencing of the simple mock metagenomic DNA using the Oxford Nanopore 16S barcoding kit SQK-RAB204 resulted in 996,441 reads following rebasecalling by albacore. These reads contained a total of 1,454,835,092 bases with an average read length of 1460 bp and a median read length of 1561 bp. 16S rRNA reads aligned by BLASTn against the Silva 16S database (version 132) with MEGAN 6 classification resulted in successful identification of three out of the 4 species, with the 4th strain being correctly identified to the genus level only (Figure 5.2A).

Rapid whole metagenome sequencing (WMGS) of the simple mock community metagenomic DNA using the SQK-RAD004 kit resulted in 97,503 reads following rebasecalling by albacore and adaptor removal. These 97,503 reads contained a total of 750,359,905 bases with an average read length of 7696 bases and a median of 5762 bases. Last alignment against the nr database followed by MEGAN long

read (LR) lowest common ancestor (LCA) analysis resulted in 74.76% bases being classified to some taxonomic level. Of these, 42.63% were classified to species level, 46.28% classified to species group level and 8.15% classified to genus level, accounting for 97.06% of classified reads. 64.37% of bases classified to genus level only were attributed to *Geobacillus*, with the remaining 35.63% classified as *Bacillus* (Figure 5.2B). Of the sequences classified to species group level, 96.32% of bases were attributed to *Bacillus cereus* group, 3.08% to *Bacillus subtilis* group, and 0.6% to *Geobacillus thermoleovrans* group. Of the sequences classified to the species level, 57.26% of bases were attributed to *Bacillus thuringiensis*, 14.74% were attributed to *Bacillus licheniformis*, 13.98% were attributed to *Bacillus cereus*, 13.8% were attributed to *Geobacillus stearothermophilus*, with only 0.21% misassigned as *Bacillus paralicheniformis* (Figure 5.2B). *De novo* assembly of raw reads from the rapid sequencing reads using the canu (version 1.7) assembler (Koren *et al.*, 2017) resulted in 104 contigs and mapping back of reads to references resulted in good coverage up to 97% identity (Figure 5.2C). The 4 reference strains contained 6 plasmids, corresponding to 10 contiguous stretches of DNA. 9 of these 10 contigs were identified following sequence assembly, the exception being pBClin15, a 15 kb plasmid from *B*. *cereus* (Figure 5.2C). 99.59% of the assembled bases aligned to the reference genomes and, of the reference genomes, 98.27% aligned to the assembled MinION sequences (Supplemental Table 5.1).

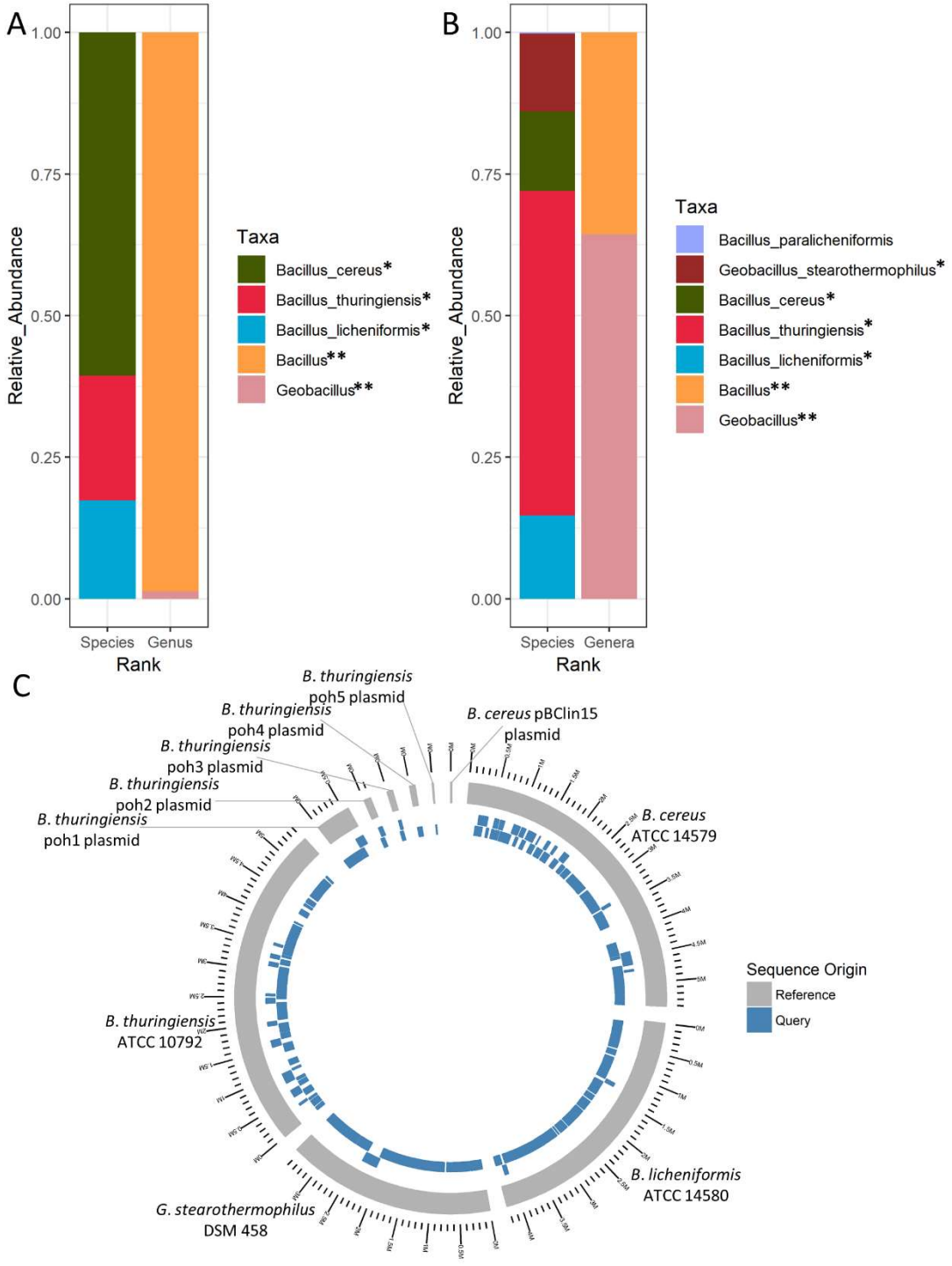


Figure 5.2. Mock community analysis.

MinION sequencing followed by MEGAN taxonomic classification of a simple mock community.

[Figure 5.2 continued]

A. Taxonomic classification following 16S sequencing. Expected species are denoted with *, whereas expected genera are denoted **.

B. Taxonomic classification following rapid WMGS. Expected species are denoted with *, whereas expected genera are denoted **.

C. *De novo* assembly of genomes by the canu assembler, followed by mapping back to original known genomes, to illustrate coverage at 97% identity. 4 genomes, with 6 plasmids illustrated, of which 4 genomes and 5 plasmids had sequences aligned at 97% identity.

5.4.2 Shotgun sequencing of environmental dairy processing samples through MinION and NextSeq sequencing provides comparable taxonomic classifications

Prompted by the successful use of MinION-based sequencing to characterise the mock metagenomic community DNA, the technology was applied to study the microbiota of a food processing facility and compared with the outputs derived through NextSeq (Illumina)-based sequencing. More specifically, eight locations in a single processing facility were swabbed over the course of a day, after CIP and before the next round of dairy processing (Figure 5.1), on three separate occasions. These eight locations comprised a table, door, wall, gaskets/flow plate seals, external surface of dryer balance tank, internal surface of dryer balance tank, external surface of evaporator, and drain beside evaporator. Overall, these eight locations were swabbed on three different months (October, November, December 2018), with one sampling day per month. These swabs were prepared for sequencing, along with a series of negative controls and a positive control, consisting of the simple mock metagenomic community used previously. Rapid sequencing of MDA DNA from 36 samples was carried out using the SQK-RBK004 rapid barcoding sequencing kit. After processing, a total of 899,306 reads were generated, containing a total of 1,648,724,928 bases with an average read length of 1,833 bases and median of 926 bases per read (and an average of 45,797,915 bases and 24,980.7 reads per sample). LAST alignment against the nr database followed by MEGAN long read (LR) lowest common ancestor (LCA) analysis resulted in 62% of bases being classified to some taxonomic level. Of these, 29.11% were classified to species level and 38.36% classified to genus level, accounting for 67.47% of

classified reads. 59 species were detected at > 5% relative abundance in at least one sample by MEGAN (Supplemental Figure 5.2).

Other shotgun sequencing-based approaches were employed to study the microbiomes of these environmental samples for comparative purposes. These included Illumina-based sequencing of MDA and non-mDNA samples, as well as of metagemomic DNA extracted from easily cultured metagenomic DNA to determine the species detected when traditional culturing-based approaches are employed. NextSeq sequencing of 93 samples produced 734,909,370 reads containing 150 bases each with an average of 7,902,251 reads per sample. To allow a comparison with MinION outputs, Diamond alignment against the nr database followed by MEGAN 6 lowest common ancestor (LCA) analysis was employed and resulted in 78% reads being classified to some taxonomic level. Of these, 10.8% were classified to the species level and 39.6% classified to the genus level, accounting for 50.3 % of classified bases. In comparison, while Kraken2 and Bracken classification resulted in 61% reads classified, to some taxonomic level, with 99% of those classified, classified to species level, this approach did not correctly classify the components of the mock metagenomic community (positive control) (supplemental Figure 5.3). Similarly, MetaPhlan did not correctly classify the mock community (Supplemental Figure 5.4), with both classifiers incorrectly classifying one species. Using the aforementioned MEGAN classification, which correctly classified the simple mock community, 108 species were identified at > 5% relative abundance in at least one sample from all MinION and NextSeq samples sequenced (Figure 5.3). Species level

classification by MEGAN revealed some agreement between corresponding NextSeq-and MinION-sequenced samples (Figure 5.3).

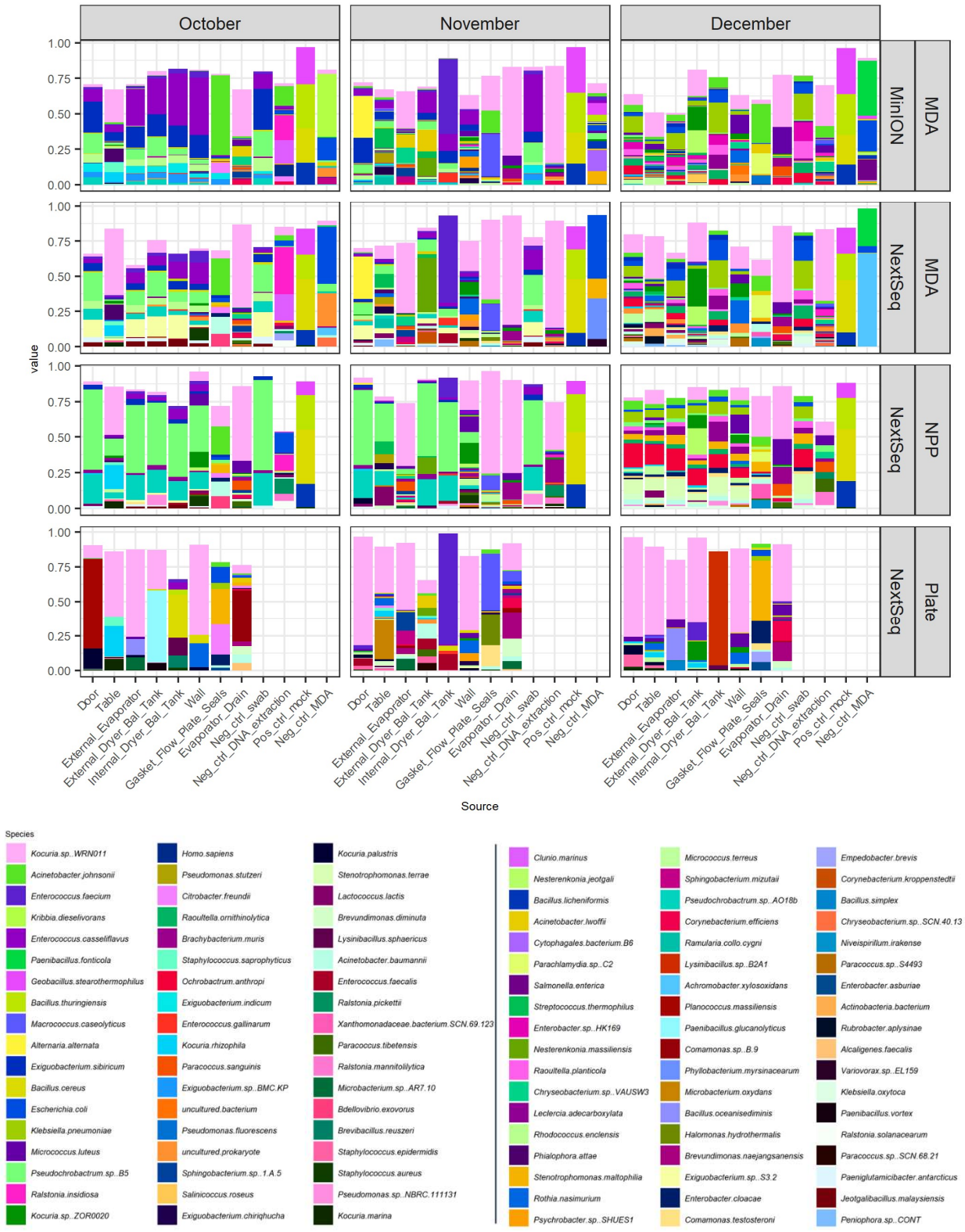


Figure 5.3. Species level classification of MinION and NextSeq sequenced environmental samples.

[Figure 5.3 continued]

Taxonomic assignment of MinION and NextSeq sequenced samples generated following the use of different pre-processing and sequencing methods. Pre-processing methods include MDA amplification, no pre-processing (NPP), and spread plating on BHI before washing colonies, pelleting, and treating as a metagenomic sample (Plate). Species level classification was performed using LAST (for MinION) and Diamond (for NextSeq) alignment of reads against the NR database and classification with MEGAN (LR for MinION). Species present in at least 5% in at least one sample are shown.

Overall, *Kocuria sp*. WRN011 was detected at the highest relative abundance. This taxon was detected in multiple locations, at each time-point, in both the MinION, and corresponding NextSeq, MDA-generated samples. Its relative abundance was highest in the evaporator drain samples at each time point. *Kocuria sp*. ZOR0020 was present in high relative abundance in external dryer balance tank swabs in both MinION- and NextSeq-MDA sequenced MDA samples (Figure 5.3). Other dominant species included *Acinetobacter johnsonnii* in gasket/flow plate seals (MinION and Illumina), *Micrococcus luteus* in evaporator drain (MinION and Illumina sequenced samples), *Enterococcus faecium* from the inside of the dryer balance tank, and many other October and November samples (MinION and MDA amplified Illumina sequencing), *Klebsiella pneumonia* in many December samples regardless of sequencing approach and *Enterococcus casseliflavus* in many samples from October and November (high relative abundance in MinION sequenced samples and at lower abundance in the corresponding MDA Illumina sequenced samples) (Figure 5.3). *Exiguobacterium sibiricum* was also detected in high relative abundance in MinION sequenced October and November door samples. It was also at lower relative abundances in many other October and November samples and in the corresponding Illumina sequenced door samples. With respect to more notable sequencing platform-dependent differences, *Exiguobacterium sp*. S3.2 and *Pseudochrobactrum sp* B5 were present at higher relative abundance in October and November MDA Illumina NextSeq sequences compared to MinION sequences, *Enterobacter sp*. HK169 was detected in December MinION samples, but not corresponding Illumina samples (Figure 5.3).

Species level taxonomic identification was performed on negative controls also. Many species identified were specific to negative controls, although some overlap with species identified in environmental samples was evident (Supplemental Document 5.1.1).

Metagenome-assembled genomes (MAGs) were extracted from assemblies of combined Illumina MDA and MinION MDA sequences. This resulted in 162 bins, of which 10 were high quality at > 80% complete and < 10% contamination (Table 5.1). 7 of the 10 MAGs were from environmental isolates, with 3 out of 10 being the positive control species used. From the remaining MAGs, 3 out of 7 environmental isolates could not be definitively assigned at the species level, being assigned as each of a number of species at similar levels of relative abundances. These MAGs were assigned at the genus level as *Planococcus*, *Exiguobacterium* and *Kocuria* and were sourced from the October evaporator drain, gasket/flow plate seal and external dryer balance tank, respectively. The MAGs that were assigned at the species level were an *Enterococcus casseliflavus* from the October table swab sample, a *Paracoccus chinensis* from the November evaporator drain, a *Micrococcus casseolyticus* from the November gasket/flow plate seal and a *Nesterenkonia massiliensis* from the November external of dryer balance tank sample (Table 5.1).

Table 5.1. High quality MAGs.

Month	Sample	Bin	Kaiju assignment	Megan assignment	Percent Complete	Percent Contamination
October	Positive control (mock)	2	*Geobacillus stearothermophilus*	*Geobacillus stearothermophilus*	91.37	1.1235
October	Evaporator Drain	5	***Planococcus plakortidis/*** *Planococcus maitriensis/ Planococcus maritimus*	***Planococcus plakortidis/*** *Planococcus maritimus/ Planococcus rifietoensis*	82.41	0.6622
October	Gasket/Flow Plate Seal	5	***Exiguobacterium acetylicum/*** *Exiguobacterium sp. RIT341/ Exiguobacterium indicum*	***Exiguobacterium indicum/*** *Exiguobacterium acetylicum*	81.9	0.6578
October	Table	14	*Enterococcus casseliflavus*	*Enterococcus casseliflavus*	82.77	1.4622
November	Positive control (mock)	1	*Bacillus licheniformis*	*Bacillus licheniformis*	94.23	0.4149
November	Evaporator Drain	7	*Paracoccus chinensis*	*Paracoccus chinensis*	93.03	1.1235
November	Gasket/Flow Plate Seal	1	*Macrococcus caseolyticus*	*Macrococcus caseolyticus*	90.35	1.1049
November	External Dryer Balance Tank	3	*Nesterenkonia massiliensis*	*Nesterenkonia massiliensis*	91.32	1.07
December	Positive control (mock)	3	*Bacillus licheniformis*	*Bacillus licheniformis*	96.29	0.0829
December	External Dryer Balance Tank	1	***Kocuria sp. CPCC 104605/*** *Kocuria sp. ZOR0020*	*Kocuria sp. ZOR0020*	81.35	0.6578

[Table 5.1 continued]

Taxonomy was assigned to metaBAT2 binned contigs by Megan LR and open reading frames of these contigs by Kaiju. If more than one species assigned, the bold species represents the top hit per classifier. Bin quality determined by checkM.

5.4.3 MDA amplification introduces bias towards the detection of some species

In order to determine the potential for bias arising from MDA pre-processing, outputs from MDA-generated NextSeq sequencing were compared to non-MDA derived NextSeq (NPP). Higher relative abundances of *Pseudochrobactrum sp*. B5 and *Pseudochrobactrum sp*. AO18b were seen in October and November NPP samples compared to the MDA-amplified equivalents (Figure 5.3). Overall, the NPP December samples we found to be less diverse than their MDA counterparts (Figure 5.3).

5.4.4 Culture-based analyses introduce a selection bias

In order to determine to what extent culture-dependent and –independent approaches provided different outputs, a comparison between NPP NextSeq-generated sequences and those resulting from sequencing of pools of easily cultured colonies (Plate samples) was performed. Sequences generated from Plate samples were noted to be considerably less diverse, although a number of the species detected were those that also been identified in the corresponding non-cultured samples (NPP and MDA amplified). This was true of *Kocuria sp* WRN011, detected in all samples in which it had previously been identified through culture-independent approaches, and *Enterococcus faecium*, the species found at highest relative abundance in all internal dryer balance tank samples from November (i.e., MDA MinION, NextSeq MDA, NPP and Plate; Figure 5.3). Pre-culturing enriched some species that had been identified at low relative abundance in metagenomic NPP and MDA samples. These included *Planococcus massiliensis* (October door sample), *Microbacterium oxydans* (November Table sample), *Acinetobacter*

baumannii (November external dryer balance tank) and *Lysinibacillus sp* B2A1 (December internal dryer balance tank; Figure 5.3).

5.4.4.1 Genus level classification highlights further culture-based selection bias

As some genera could not be distinguished at the level, genus level assignments were also investigated and compared. MEGAN LCA analysis identified sequences that couldn't be more accurately classified to species level, and assigned these as far as genus level only. A combined fifty-six genera were identified between MinION, NextSeq (both at > 5% relative abundance) and Sanger sequencing. Fifteen of these 56 genera were identified in samples from all 3 sequencing types (Supplemental Figure 5.5). Sanger sequencing involved partial 16S rRNA sequencing of morphologically different colonies from BHI plates, including total spread plate (TBC), thermophilic enriched spore pasteurised (ST) and mesophilic enriched spore pasteurised (SM) tests (Supplemental Table 5.2; Figure 5.4). There was agreement between Sanger sequencing of isolates and next generation sequencing of plate samples with respect to *Kocuria*, *Acinetobacter* and *Lysinibacillus* (Figure 5.4). Some genera identified in Plate NextSeq samples and Sanger sequences had not been seen in high relative abundance in corresponding culture-independent NextSeq or MinION sequencing. These include *Microbacterium* in the November table sample and *Lysinibacillus* in the December internal dryer balance tank (Figure 5.4).

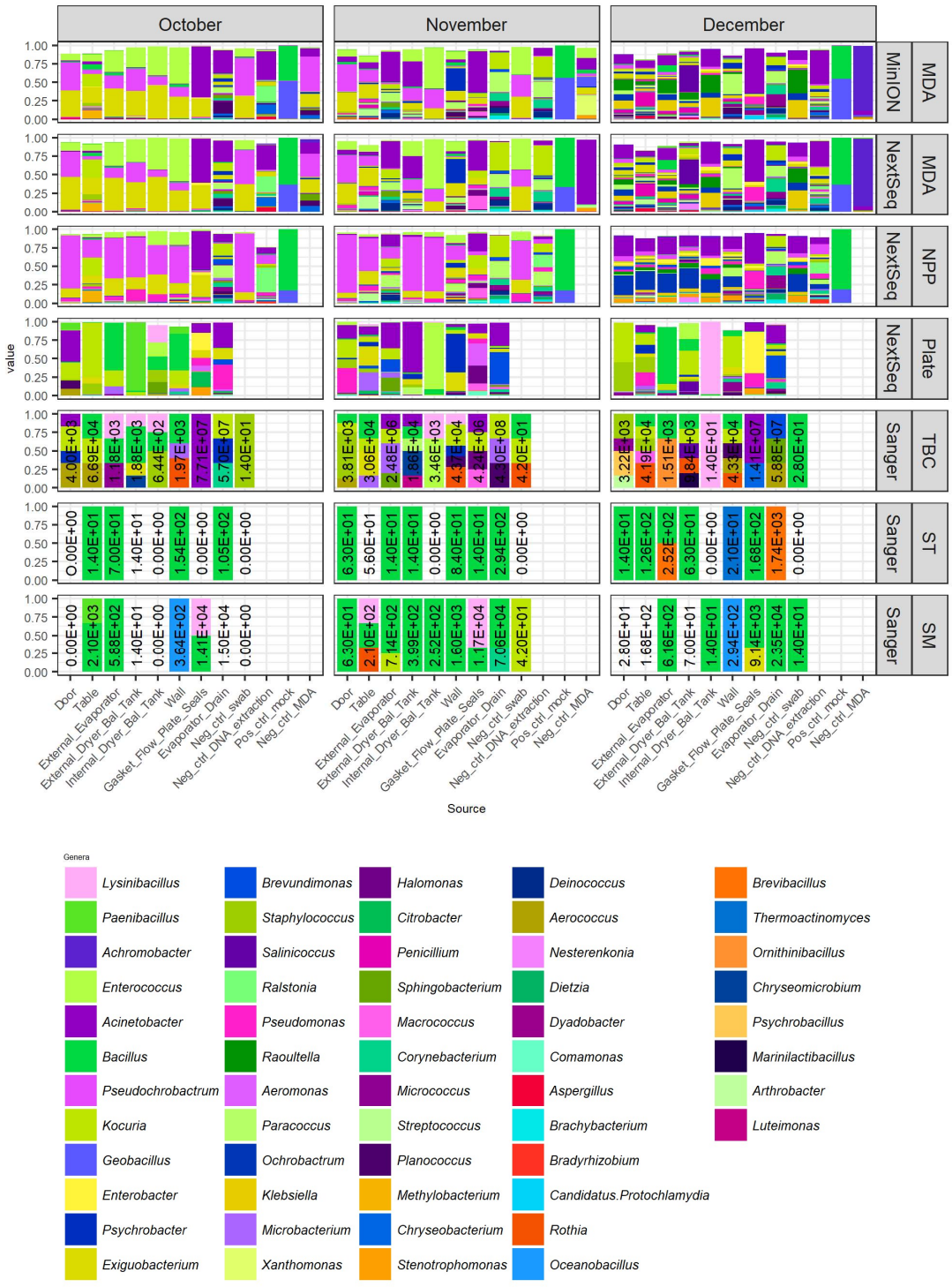


Figure 5.4. Genus level classification of environmental samples and controls following different pre-processing methods and sequencing methods.

MEGAN LCA based genera level classification of MinION and NextSeq sequences.

[Figure 5.4 continued]

Also depicted are Sanger results to genus level for morphologically different colonies from each sample (TBC) along with thermophilic sporeformer enriched (ST) and mesophilic sporeformer enriched (SM) counts. Also included are CFU / swab counts for each culturing type. Sanger results represent relative abundance of a subset of morphologically distinct isolates rather than total isolates.

Overall, Sanger sequencing of 16S variable region of TBC isolates corresponded well with NextSeq 'Plate' sequencing but fewer genera were identified per sample. This may in part be due to only very morphologically distinct isolates being selected for Sanger sequencing. Counts per swab are also included. At all timepoints the gasket/flow plate seals and the evaporator drains had highest CFU / swab, with on average 3.18 x 10^7 CFU / swab and 1.82 x 10^8 CFU / swab each. These two areas also had the highest mesophilic spore count with an average of 1.17 x 10^4 CFU / swab and 3.64 x10^4 CFU / swab each (Figure 5.4, Supplemental Table 5.2).

5.4.5 Relatively few significant differences in relative abundance of species and genus level taxonomic classification due to sequencing and pre-processing approaches

Overall, only six out of 108 species had significantly different relative abundance between environmental samples (excluding controls) due to sample processing or sequencing method, based on Pairwise Wilcoxon rank sums test using Benjamini Hochberg *p*-value correction analysis of sequential pairs (Supplemental Figure 5.6). *Enterococcus casseliflavus*, *Acinetobacter lwoffii* and *Acinetobacter johnsonii* had significantly higher relative abundance in MDA MinION sequenced samples than MDA NextSeq sequenced samples, whereas *Kocuria sp*. WRN011 was identified at significantly higher relative abundance in MDA NextSeq samples than MDA MinION samples. *Pseudochrobactrum sp* B5 was detected at significantly higher relative abundance in NPP NextSeq samples than MDA processed NextSeq samples, whereas *Exiguobacterium sibricum* was detected at significantly higher relative

abundance in MDA NextSeq samples compared to NPP NextSeq samples (Figure 5.5).

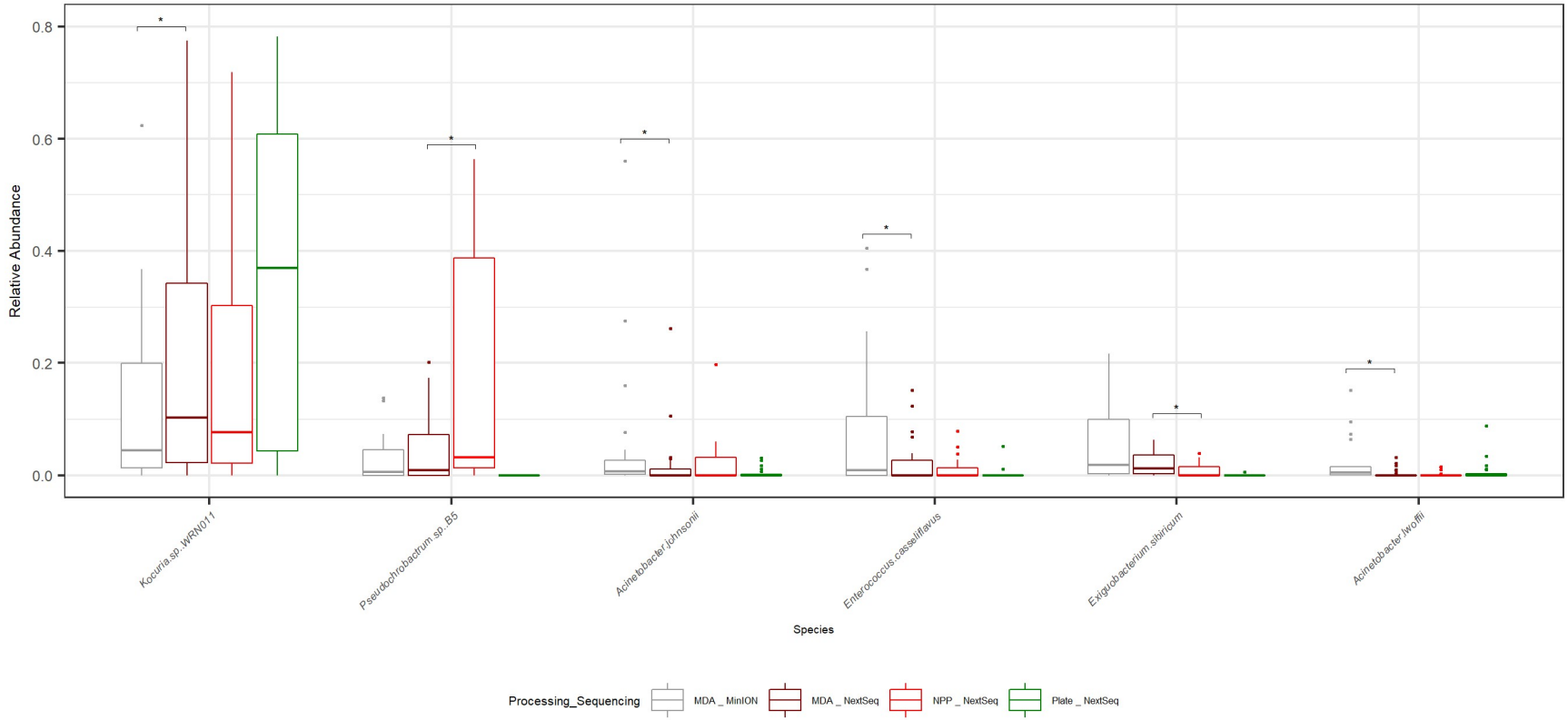


Figure 5.5. Significant species level differences between sequencing and processing types on environmental samples.

Significant differences in the relative abundance of species base on processing and sequencing type. Controls are excluded from these calculations and plots.

Genera that had significantly different relative abundances, depending on whether MinION or NextSeq sequencing approaches were used, were also identified. In this case a greater number of significantly different taxa were noted, with 24 genera out of a total of 46 being significantly different as a consequence of the sample processing or sequencing method used (Figure 5.6, Supplemental Figure 5.7). Six genera are seen to differ significantly between more than one pairwise group (Figure 5.6). *Pseudochrobactrum* was present at significantly different relative abundances across all 3 pairwise groups (i.e., MDA MinION and MDA NextSeq, MDA NextSeq and NPP NextSeq, and NPP NextSeq and Plate NextSeq). *Exiguobacterium* and *Planococcus* were present at significantly different relative abundances between MDA MinION and MDA NextSeq as well as MDA NextSeq and NPP NextSeq. *Bacillus*, *Staphylococcus* and *Ochrobactrum* were present at significantly different relative abundances between MDA NextSeq and NPP NextSeq as well as NPP NextSeq and Plate NextSeq. The remaining 18 genera only differed across one pair of analyses (Figure 5.6).

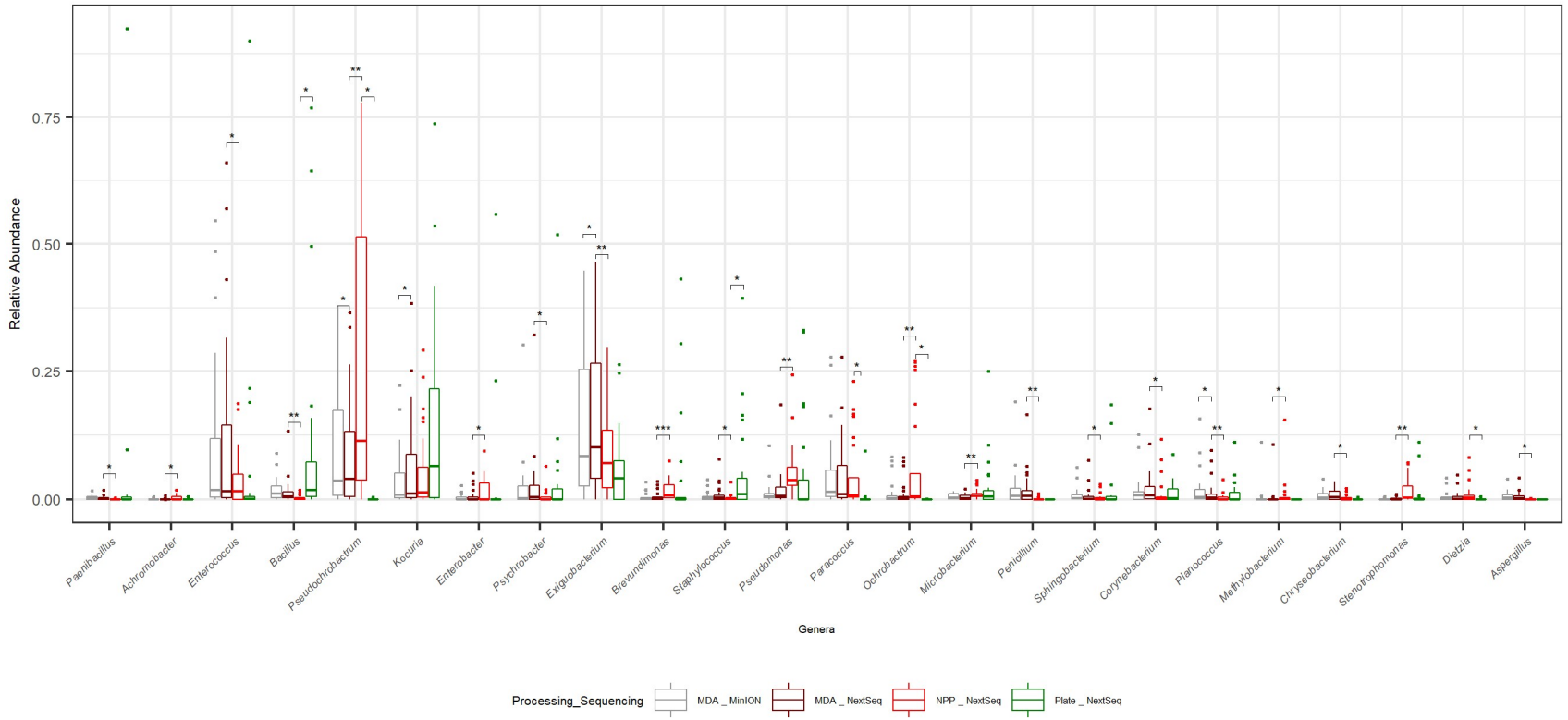


Figure 5.6. Significant differences at genera level of environmental samples by pre-processing type and sequencing method.

Significant differences in the relative abundance of genera level LCA classification based on processing and sequencing type. Controls are excluded from these calculations and plot.

5.5 Discussion

16S rRNA rapid barcoding-based MinION sequencing of a simple mock community coupled with MEGAN classification by aligning with BLAST against a Silva database provided species level classification to 3 of the 4 species in a mock community and correctly identified both genera present. The rapid sequencing kit-based shotgun sequencing on the MinION platform coupled with LAST alignment against NR database and MEGAN taxonomic classification resulted in correct classification of all four species but with low level false positive detection of *B*. *paralicheniformis*, a close relative of *B. licheniformis*. Thus, in this regard, MinION rapid WMGS performed better than MinION 16S sequencing for species level classification of related species, and could be further improved by reducing/eliminating false positives by exercising a stricter cut off and only focusing on species detected at high relative abundance.

Environmental DNA samples subject to MDA resulted in MinION sequencing reads that were shorter, with lower output than the high quality, high quantity, pooled mock metagenomic DNA generated. This is a particular issue for sequencing of low biomass environmental samples where, without the use of MDA, the quantities of DNA would not suffice for current rapid protocols, even after pooling of multiple swabs. The multiplexing of poorer quality DNA from environmental samples resulted in saturation of flow cells, resulting in lower output compared to the mock sequencing run. Despite this, MinION sequencing of environmental sample did perform well and was comparable to other methods when all factors were considered. Some of the most abundant species identified included *Kocuria* sp.

WRN011, *Enterococcus casseliflavus* and *Enterococcus faecium*. *Kocuria* sp. WRN011 is a saline alkaline soil isolate, and is perhaps selected for due to the unfavourable conditions within a food processing environment arising after cleaning in place (CIP). Both *Enterococcus faecium* and *Enterococcus casseliflavus* are common dairy microorganisms (Rivas *et al.*, 2012; Gelsomino *et al.*, 2002), with *Enterococcus* sp. been known to also be capable of growth at high pH and in the presence of NaCl (Khedid *et al.*, 2009).

More than one species was assigned to each of twenty four genera classified. These species were often co-localised, accounting for 68 out of 108 species (Supplemental Document 5.1.2), and thus could be due an inability to distinguish between these species. MAG analysis revealed 10 good quality genomes from combined MinION and MDA Illumina sequence reads. Seven of the 10 genomes originated from environmental swab samples, with the other 3 corresponding to positive controls. This form of analysis can, if carried out on a larger scale in the future and with greater sequencing depth, be used to bridge discrepancies in taxonomic classification.

There were also significant differences in the relative abundance of species due to the pre-processing and sequencing approaches taken. MinION sequencing indicated greater relative abundances of *Enterococcus casseliflavus*, *Acinetobacter lwoffii* and *Acinteobacter johnsonnii* than was suggested by MDA NextSeq sequencing. NextSeq MDA appeared to preferentially sequence *Kocuria sp*. WRN011 compared to MinION. *Pseudochrobactrum sp*. B5 abundances appeared lower in MDA (MinION and NextSeq) and easily culturable NextSeq plate samples than NPP samples. From

a culture-based perspective, it is noted that this species is known to reduce hexavalent chromium (Ge, Dong and Zhou, 2013) and it may not grow well on the BHI agar used. *Exiguobacterium sibricum* was detected in higher relative abundance in MDA amplified samples, with significantly higher relative abundance in MDA NextSeq samples compared to NPP NextSeq samples. This suggests it is preferentially amplified by multiple displacement amplification, leading to an overestimation of its relative abundance in these samples.

There were also significant differences in the relative abundances of genera that could not be assigned at the species level. This was most apparent when MDA NextSeq and NPP NextSeq outputs were compared. As well as plate sequences having lower levels of *Pseudochrobactrum*, they were also a lot less diverse than those generated through culture-independent approaches, suggesting culturing at the conditions used was less sensitive. Many species were seen in higher relative abundance in NextSeq plate samples than samples not subject to pre-culturing, including *Planococcus massiliensis*, *Microbacterium oxydans*, *Acinetobacter baumannii* and *Lysinibacillus sp.* B2A1, presumably as a consequence of being better suited to growth in these conditions.

Small, portable, real-time DNA sequencers provide the first steps towards real-time industry paced microbial classification and analysis, which could allow the implementation of process change to counteract microbial issues. Although DNA sequencing has been used sporadically for source tracking (Doyle *et al.*, 2017; Fretin *et al.*, 2018) and monitoring the microbiota through various seasons and environmental conditions (Li *et al.*, 2018), there are currently limited numbers of

publications and datasets relating to food chain and processing facility microbiomes. While Oxford Nanopore sequencing accuracy is constantly improving (Watson and Warr, 2019), this in itself provides another hurdle to routine implementation in food processing environments, due to often lack of back compatibility with kits, hardware, software and analysis pipelines. More importantly, the need for high quality, high quantity DNA from swabs of an area that aims to have low bacterial loads is a challenge. Ideally, future forms of portable technologies can be implemented with a rapid kit, without a need for amplification. Despite these challenges, this study and the data generated will aid further attempts to characterise the microbiotas across the food chain, leading to an acceleration towards routine implementation. This is particularly true regarding the generation of MAGs from MDA amplified DNA, resulting in good quality MAGs for 7 environmental isolates, for which relatively few genomes are already available. Notably, in some cases it was difficult to assign some of these MAGs to an existing species, suggesting that the genomes isolated were from related, but previously unclassified, species. While *Exiguobacterium sp*. and *Kocuria sp*. have previously been reported in food processing environments (Vishnivetskaya and Kathariou, 2005; Røder *et al.*, 2015), *Planococcus sp*., although not well characterised with few genomes available, are regarded as halotolerant, water-associated microorganisms, rather than food processing contaminants (Waghmode *et al.*, 2019). The generation of this MAG and further generation of MAGs, will accelerate the identification of food chain microbes through sequencing-based approaches in the future.

5.6 Conclusion

Ultimately, while this study highlights issues relating to sourcing sufficient template DNA, inconsistencies across sequencing approaches and platforms, and challenges with assigning taxa, the considerably great potential merits of applying shotgun metagenomic approaches to monitor the microbiology of the food chain are clear.

5.7 References

Altschul, S., Gish, W., Miller, W., Myers, E. and Lipman, D. (1990) 'Basic local alignment search tool', *J Mol Biol,* 215. DOI: 10.1016/s0022-2836(05)80360-2

Bertrand, D., Shaw, J., Kalathiyappan, M., Ng, A. H. Q., Kumar, M. S., Li, C., Dvornicic, M., Soldo, J. P., Koh, J. Y., Tong, C., Ng, O. T., Barkham, T., Young, B., Marimuthu, K., Chng, K. R., Sikic, M. and Nagarajan, N. (2019) 'Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes', *Nat Biotechnol,* 37(8), pp. 937-944. DOI: 10.1038/s41587-019-0191-2

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics (Oxford, England),* 30(15), pp. 2114-2120. DOI: 10.1093/bioinformatics/btu170

Buchfink, B., Xie, C. and Huson, D. H. (2015) 'Fast and sensitive protein alignment using DIAMOND', *Nat Meth,* 12(1), pp. 59-60. DOI: 10.1038/nmeth.3176

Burgess, S. A., Lindsay, D. and Flint, S. H. (2010) 'Thermophilic bacilli and their importance in dairy processing', *Int J Food Microbiol,* 144(2), pp. 215-25. DOI: 10.1016/j.ijfoodmicro.2010.09.027

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M.-A., Barrell, B. G. and Parkhill, J. (2005) 'ACT: the Artemis comparison tool', *Bioinformatics,* 21(16), pp. 3422-3423. DOI: 10.1093/bioinformatics/bti553

Charalampous, T., Kay, G. L., Richardson, H., Aydin, A., Baldan, R., Jeanes, C., Rae, D., Grundy, S., Turner, D. J., Wain, J., Leggett, R. M., Livermore, D. M. and O'Grady,

J. (2019) 'Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection', *Nat Biotechnol*. DOI: 10.1038/s41587-019-0156-5

Cho, T. J., Kim, H. W., Kim, N. H., Park, S. M., Kwon, J. I., Kim, Y. J., Lee, K. W. and Rhee, M. S. (2018) 'New insights into the thermophilic spore-formers in powdered infant formula: Implications of changes in microbial composition during manufacture', *Food Control,* 92, pp. 464-470. DOI: 10.1016/j.foodcont.2018.05.036

Delcher, A. L., Phillippy, A., Carlton, J. and Salzberg, S. L. (2002) 'Fast algorithms for large-scale genome alignment and comparison', *Nucleic acids research,* 30(11), pp. 2478-2483. DOI: 10.1093/nar/30.11.2478

Doyle, C. J., Gleeson, D., Jordan, K., Beresford, T. P., Ross, R. P., Fitzgerald, G. F. and Cotter, P. D. (2015) 'Anaerobic sporeformers and their significance with respect to milk and dairy products', *Int J Food Microbiol,* 197, pp. 77-87. DOI: 10.1016/j.ijfoodmicro.2014.12.022

Doyle, C. J., Gleeson, D., O'Toole, P. W. and Cotter, P. D. (2017) 'Impacts of seasonal housing and teat preparation on raw milk microbiota: a high-throughput sequencing study', *Appl Environ Microbiol,* 83(2). DOI: 10.1128/AEM.02694-16

Doyle, C. J., O'Toole, P. W. and Cotter, P. D. (2018) 'Genomic Characterization of Sulphite Reducing Bacteria Isolated From the Dairy Production Chain', *Frontiers in Microbiology,* 9. DOI: 10.3389/fmicb.2018.01507

Faille, C., Benezech, T., Midelet-Bourdin, G., Lequette, Y., Clarisse, M., Ronse, G., Ronse, A. and Slomianny, C. (2014) 'Sporulation of Bacillus spp. within biofilms: a

potential source of contamination in food processing environments', *Food Microbiol,* 40, pp. 64-74. DOI: 10.1016/j.fm.2013.12.004

Fretin, M., Martin, B., Rifa, E., Isabelle, V. M., Pomies, D., Ferlay, A., Montel, M. C. and Delbes, C. (2018) 'Bacterial community assembly from cow teat skin to ripened cheeses is influenced by grazing systems', *Sci Rep,* 8(1), pp. 200. DOI: 10.1038/s41598-017-18447-y

Fysun, O., Kern, H., Wilke, B. and Langowski, H.-C. (2019) 'Evaluation of factors influencing dairy biofilm formation in filling hoses of food-processing equipment', *Food and Bioproducts Processing,* 113, pp. 39-48. DOI: 10.1016/j.fbp.2018.10.009

Ge, S., Dong, X. and Zhou, J. (2013) 'Comparative evaluations on bio-treatment of hexavalent chromate by resting cells of Pseudochrobactrum sp. and Proteus sp. in wastewater', *J Environ Manage,* 126, pp. 7-12. DOI: 10.1016/j.jenvman.2013.04.011

Gelsomino, R., Vancanneyt, M., Cogan, T. M., Condon, S. and Swings, J. (2002) 'Source of Enterococci in a Farmhouse Raw-Milk Cheese', *Applied and Environmental Microbiology,* 68(7), pp. 3560. DOI: 10.1128/aem.68.7.3560-3565.2002

Gleeson, D., O'Connell, A. and Jordan, K. (2013) 'Review of potential sources and control of thermoduric bacteria in bulk-tank milk', *Irish Journal of Agricultural and Food Research,* 52(2), pp. 217-227.

Huson, D., Auch, A., Qi, J. and Schuster, S. (2007) 'MEGAN analysis of metagenomic data', *Genome Res,* 17. DOI: 10.1101/gr.5969107

Huson, D. H., Albrecht, B., Bagci, C., Bessarab, I., Gorska, A., Jolic, D. and Williams, R. B. H. (2018) 'MEGAN-LR: new algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs', *Biol Direct,* 13(1), pp. 6. DOI: 10.1186/s13062-018-0208-7

Kafetzopoulou, L. E., Efthymiadis, K., Lewandowski, K., Crook, A., Carter, D., Osborne, J., Aarons, E., Hewson, R., Hiscox, J. A., Carroll, M. W., Vipond, R. and Pullan, S. T. (2018) 'Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples', *Euro Surveill,* 23(50). DOI: 10.2807/1560-7917.es.2018.23.50.1800228

Kang, D. D., Froula, J., Egan, R. and Wang, Z. (2015) 'MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities', *PeerJ,* 3, pp. e1165. DOI: 10.7717/peerj.1165

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H. and Wang, Z. (2019) 'MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies', *PeerJ,* 7, pp. e7359-e7359. DOI: 10.7717/peerj.7359

Khedid, K., Faid, M., Mokhtari, A., Soulaymani, A. and Zinedine, A. (2009) 'Characterization of lactic acid bacteria isolated from the one humped camel milk produced in Morocco', *Microbiological Research,* 164(1), pp. 81-91. DOI: 10.1016/j.micres.2006.10.008

Kielbasa, S. M., Wan, R., Sato, K., Horton, P. and Frith, M. C. (2011) 'Adaptive seeds tame genomic sequence comparison', *Genome Res,* 21(3), pp. 487-93. DOI: 10.1101/gr.113985.110

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H. and Phillippy, A. M. (2017) 'Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation', *Genome Res,* 27(5), pp. 722-736. DOI: 10.1101/gr.215087.116

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C. and Salzberg, S. L. (2004) 'Versatile and open software for comparing large genomes', *Genome biology,* 5(2), pp. R12-R12. DOI: 10.1186/gb-2004-5-2-r12

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nat Meth,* 9(4), pp. 357-359. DOI: 10.1038/nmeth.1923

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T. and Carey, V. J. (2013) 'Software for Computing and Annotating Genomic Ranges', *PLOS Computational Biology,* 9(8), pp. e1003118. DOI: 10.1371/journal.pcbi.1003118

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics,* 25(16), pp. 2078-9. DOI: 10.1093/bioinformatics/btp352

Li, N., Wang, Y., You, C., Ren, J., Chen, W., Zheng, H. and Liu, Z. (2018) 'Variation in Raw Milk Microbiota Throughout 12 Months and the Impact of Weather Conditions', *Sci Rep,* 8(1), pp. 2371. DOI: 10.1038/s41598-018-20862-8

Lu, J., Breitwieser, F. P., Thielen, P. and Salzberg, S. L. (2017) 'Bracken: estimating species abundance in metagenomics data', *PeerJ Computer Science,* 3, pp. e104. DOI: 10.7717/peerj-cs.104

McHugh, A. J., Feehily, C., Tobin, J. T., Fenelon, M. A., Hill, C. and Cotter, P. D. (2018) 'Mesophilic Sporeformers Identified in Whey Powder by Using Shotgun Metagenomic Sequencing', *Appl Environ Microbiol,* 84(20). DOI: 10.1128/AEM.01305-18

McIver, L. J., Abu-Ali, G., Franzosa, E. A., Schwager, R., Morgan, X. C., Waldron, L., Segata, N. and Huttenhower, C. (2018) 'bioBakery: a meta'omic analysis environment', *Bioinformatics (Oxford, England),* 34(7), pp. 1235-1237. DOI: 10.1093/bioinformatics/btx754

Menzel, P., Ng, K. L. and Krogh, A. (2016) 'Fast and sensitive taxonomic classification for metagenomics with Kaiju', *Nat Commun,* 7, pp. 11257. DOI: 10.1038/ncomms11257

Peng, Y., Leung, H. C. M., Yiu, S. M. and Chin, F. Y. L. (2012) 'IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth', *Bioinformatics,* 28(11), pp. 1420-1428. DOI: 10.1093/bioinformatics/bts174

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J. and Glöckner, F. O. (2007) 'SILVA: a comprehensive online resource for quality checked and

aligned ribosomal RNA sequence data compatible with ARB', *Nucleic Acids Research,* 35(21), pp. 7188-7196. DOI: 10.1093/nar/gkm864

Pruitt, K. D., Tatusova, T., Brown, G. R. and Maglott, D. R. (2012) 'NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy', *Nucleic Acids Res,* 40. DOI: 10.1093/nar/gkr1079

Pruitt, K. D., Tatusova, T. and Maglott, D. R. (2005) 'NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins', *Nucleic acids research,* 33(Database issue), pp. D501-D504. DOI: 10.1093/nar/gki025

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F. O. (2012) 'The SILVA ribosomal RNA gene database project: improved data processing and web-based tools', *Nucleic Acids Research,* 41(D1), pp. D590-D596. DOI: 10.1093/nar/gks1219

Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T. F., Beutler, N. A., Burton, D. R., Lewis-Ximenez, L. L., de Jesus, J. G., Giovanetti, M., Hill, S. C., Black, A., Bedford, T., Carroll, M. W., Nunes, M., Alcantara Jr, L. C., Sabino, E. C., Baylis, S. A., Faria, N. R., Loose, M., Simpson, J. T., Pybus, O. G., Andersen, K. G. and Loman, N. J. (2017) 'Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples', *Nature Protocols,* 12, pp. 1261. DOI: 10.1038/nprot.2017.066

Rivas, F. P., Castro, M. P., Vallejo, M., Marguet, E. and Campos, C. A. (2012) 'Antibacterial potential of Enterococcus faecium strains isolated from ewes' milk

and cheese', *LWT - Food Science and Technology,* 46(2), pp. 428-436. DOI: 10.1016/j.lwt.2011.12.005

Røder, H. L., Raghupathi, P. K., Herschend, J., Brejnrod, A., Knøchel, S., Sørensen, S. J. and Burmølle, M. (2015) 'Interspecies interactions result in enhanced biofilm formation by co-cultures of bacteria isolated from a food processing environment', *Food Microbiology,* 51, pp. 18-24. DOI: 10.1016/j.fm.2015.04.008

Sadiq, F. A., Li, Y., Liu, T., Flint, S., Zhang, G., Yuan, L., Pei, Z. and He, G. (2016) 'The heat resistance and spoilage potential of aerobic mesophilic and thermophilic spore forming bacteria isolated from Chinese milk powders', *Int J Food Microbiol,* 238, pp. 193-201. DOI: 10.1016/j.ijfoodmicro.2016.09.009

Sanderson, N. D., Street, T. L., Foster, D., Swann, J., Atkins, B. L., Brent, A. J., McNally, M. A., Oakley, S., Taylor, A., Peto, T. E. A., Crook, D. W. and Eyre, D. W. (2018) 'Real-time analysis of nanopore-based metagenomic sequencing from infected orthopaedic devices', *BMC Genomics,* 19(1), pp. 714. DOI: 10.1186/s12864-018-5094-y

Seemann, T. (2014) 'Prokka: rapid prokaryotic genome annotation', *Bioinformatics,* 30(14), pp. 2068-9. DOI: 10.1093/bioinformatics/btu153

Sheetlin, S. L., Park, Y., Frith, M. C. and Spouge, J. L. (2014) 'Frameshift alignment: statistics and post-genomic applications', *Bioinformatics (Oxford, England),* 30(24), pp. 3575-3582. DOI: 10.1093/bioinformatics/btu576

Tallent, S. M., Kotewicz, K. M., Strain, E. A. and Bennett, R. W. (2012) 'Efficient isolation and identification of Bacillus cereus group', *J AOAC Int,* 95(2), pp. 446-51. DOI: 10.5740/jaoacint.11-251

Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C. and Segata, N. (2015) 'MetaPhlAn2 for enhanced metagenomic taxonomic profiling', *Nat Meth,* 12(10), pp. 902-903. DOI: 10.1038/nmeth.3589

Vishnivetskaya, T. A. and Kathariou, S. (2005) 'Putative transposases conserved in Exiguobacterium isolates from ancient Siberian permafrost and from contemporary surface habitats', *Applied and environmental microbiology,* 71(11), pp. 6954-6962. DOI: 10.1128/aem.71.11.6954-6962.2005

Waghmode, S., Suryavanshi, M., Dama, L., Kansara, S., Ghattargi, V., Das, P., Banpurkar, A. and Satpute, S. K. (2019) 'Genomic Insights of Halophilic Planococcus maritimus SAMP MCC 3013 and Detail Investigation of Its Biosurfactant Production', *Frontiers in microbiology,* 10, pp. 235-235. DOI: 10.3389/fmicb.2019.00235

Wang, B., Tan, X., Du, R., Zhao, F., Zhang, L., Han, Y. and Zhou, Z. (2019) 'Bacterial composition of biofilms formed on dairy-processing equipment', *Preparative Biochemistry and Biotechnology,* 49(5), pp. 477-484. DOI: 10.1080/10826068.2019.1587623

Watson, M. and Warr, A. (2019) 'Errors in long-read assemblies can critically affect protein prediction', *Nature Biotechnology,* 37(2), pp. 124-126. DOI: 10.1038/s41587-018-0004-z

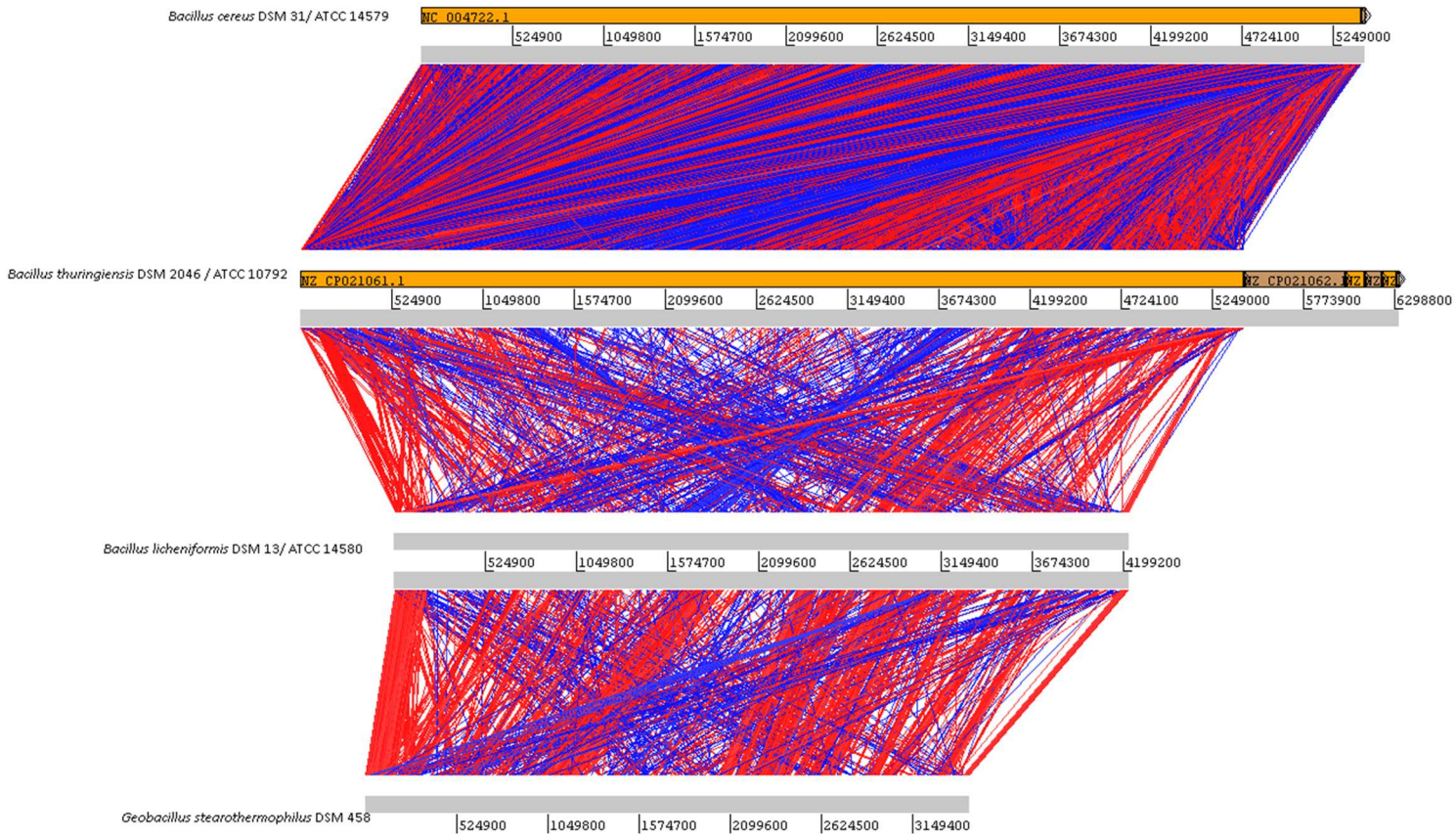
Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis.* Springer Publishing Company, Incorporated. DOI: 10.1007/978-0-387-98141-3

Wood, D. E. and Salzberg, S. L. (2014) 'Kraken: ultrafast metagenomic sequence classification using exact alignments', *Genome Biology,* 15(3), pp. R46. DOI: 10.1186/gb-2014-15-3-r46

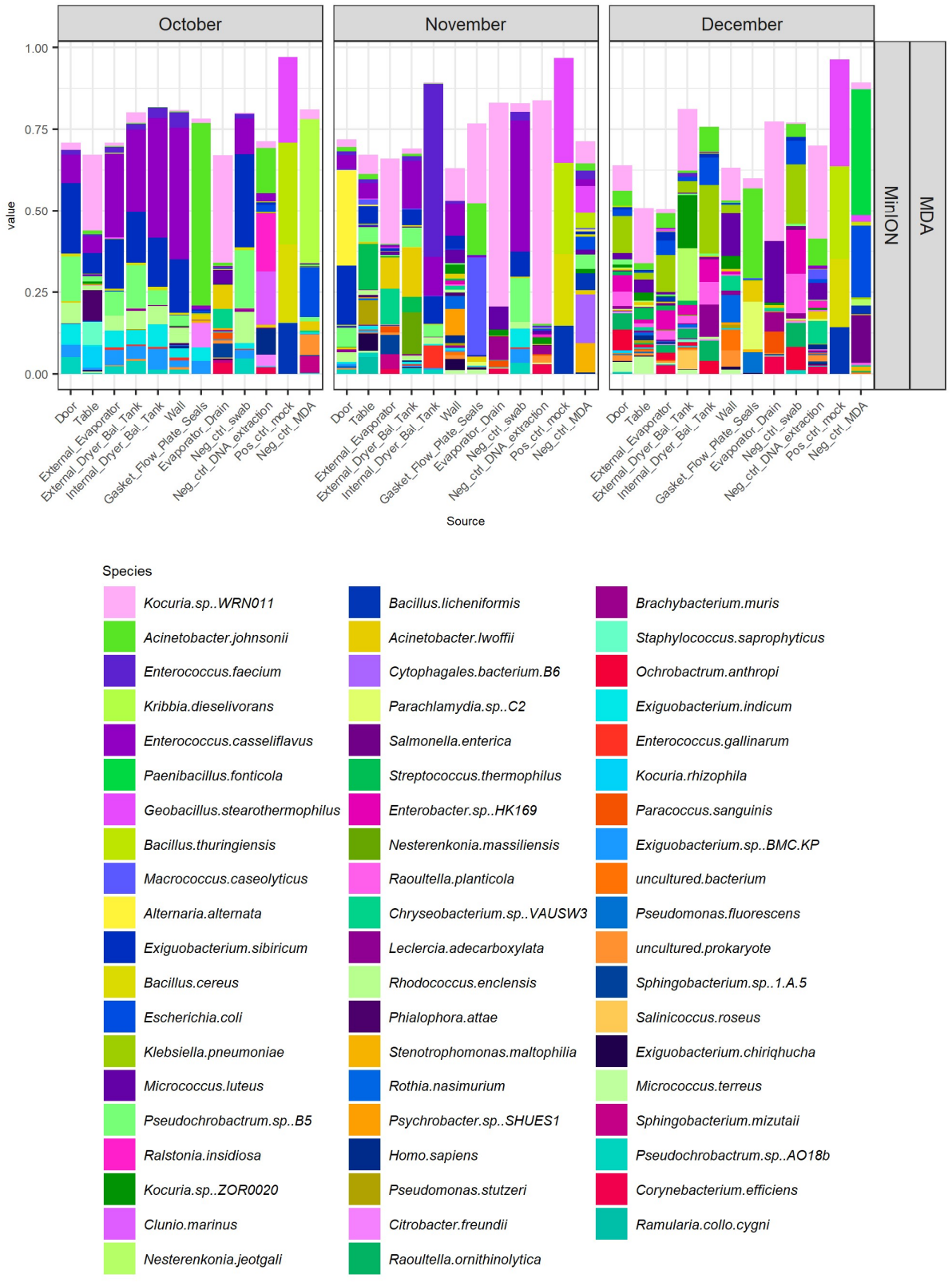
Yin, T., Cook, D. and Lawrence, M. (2012) 'ggbio: an R package for extending the grammar of graphics for genomic data', *Genome Biology,* 13(8), pp. R77. DOI: 10.1186/gb-2012-13-8-r77

Website 1 (2018). https://github.com/ms-gx/guppy_bcsplit/blob/master/guppy_bcsplit.py. Last accessed 04/03/2020.

5.8 Supplemental Figures



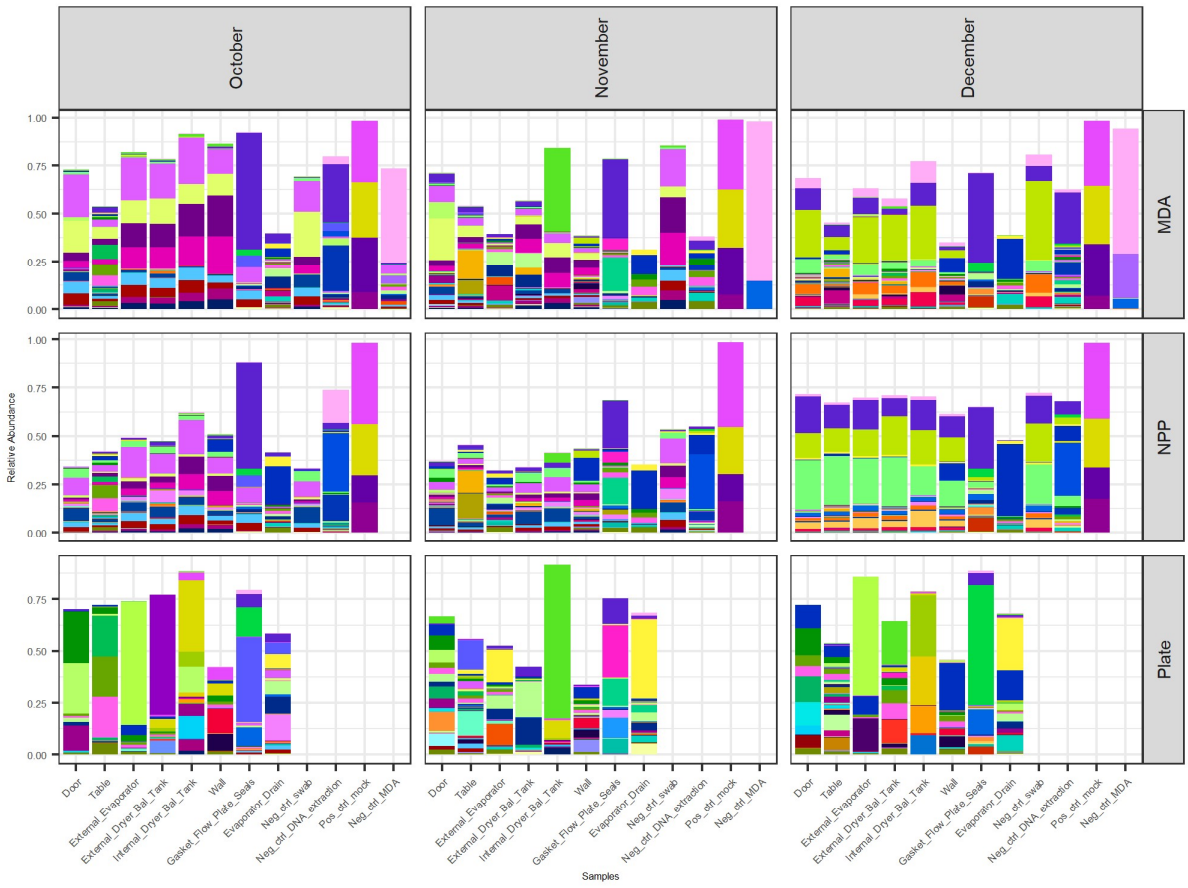
Supplemental Figure 5.1. Comparison of genomes of 4 strains used in mock community using ACT.



Supplemental Figure 5.2. Species level classification of MinION environmental samples.

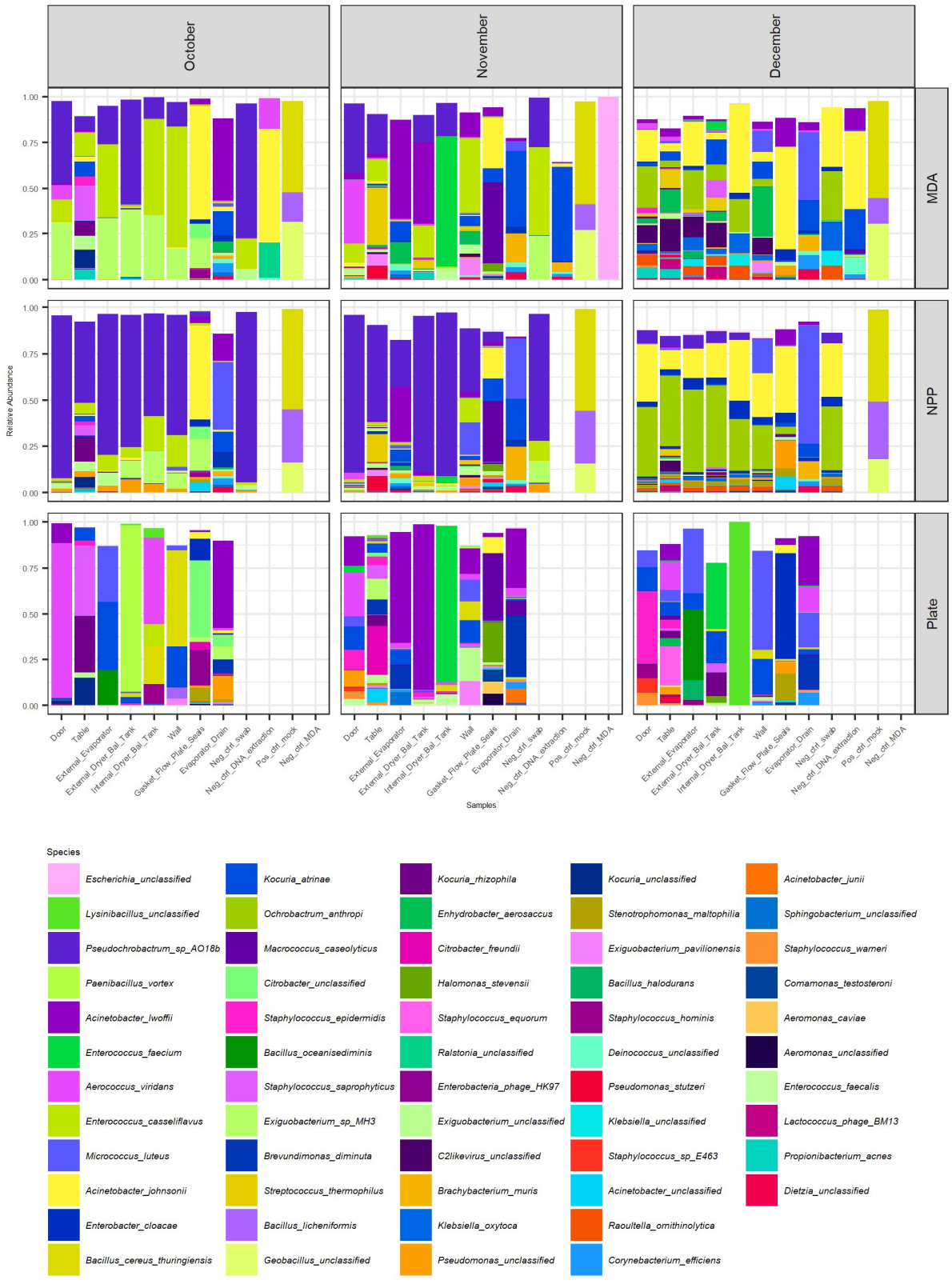
[Supplemental Figure 5.2 continued]

Oxford nanopore MinION sequencing of MDA DNA from environmental swab samples classified using LAST+MEGAN LR. Species present > 5% relative abundance in at least one sample are shown.

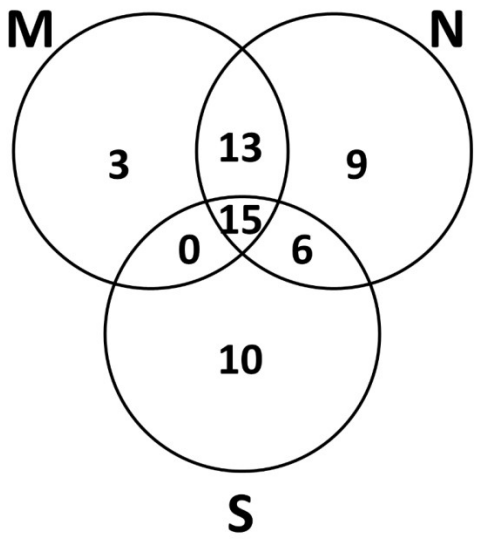


Supplemental Figure 5.3. Bracken classification of NextSeq sequences.

Kraken2 with Bracken species level classification of NextSeq samples. 100% mock community was classified, but yet couldn't accurately decipher species level classification.

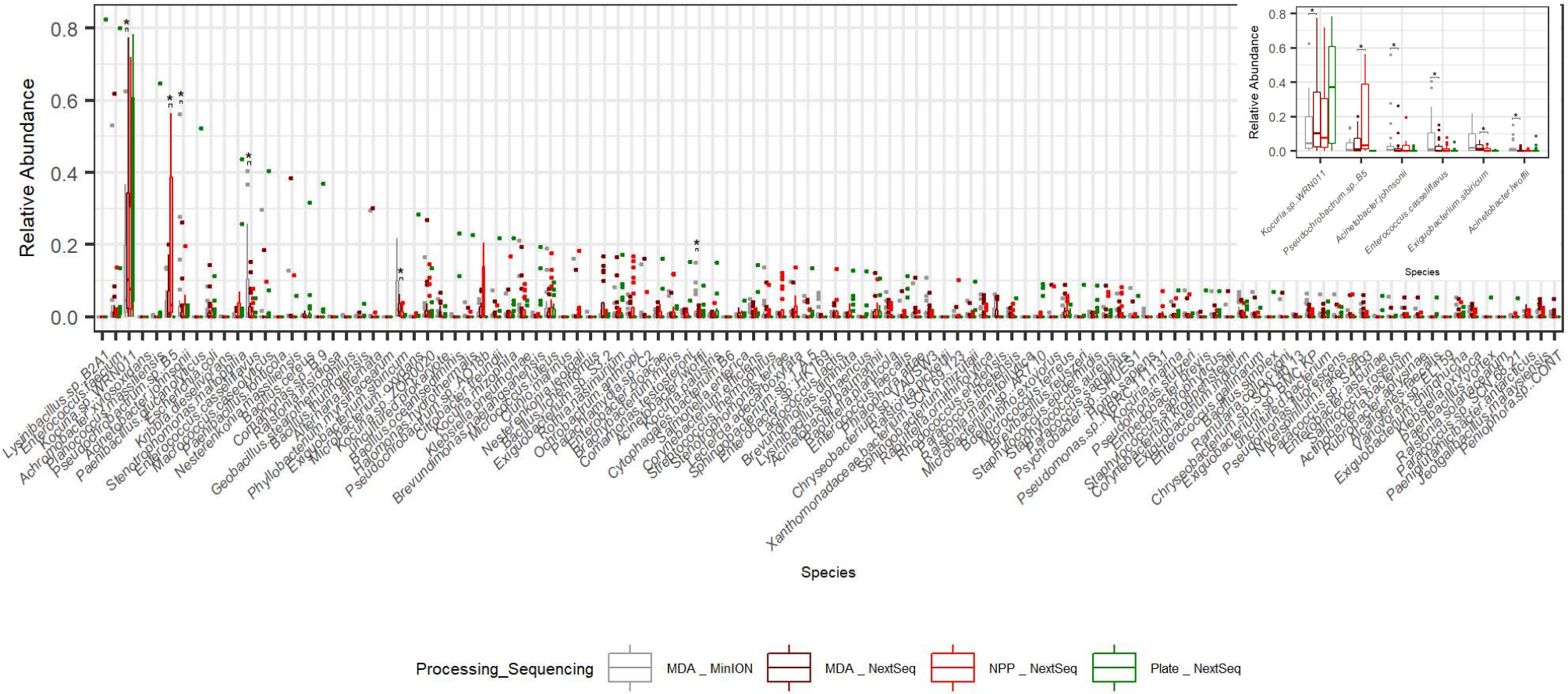


Supplemental Figure 5.4. Metaphlan species level classification of NextSeq data. Metaphlan species level classification of NextSeq data. Again highlights an incorrectly classified positive control/mock community.



Supplemental Figure 5.5. Venn diagram of the number of genera assigned per sequence type.

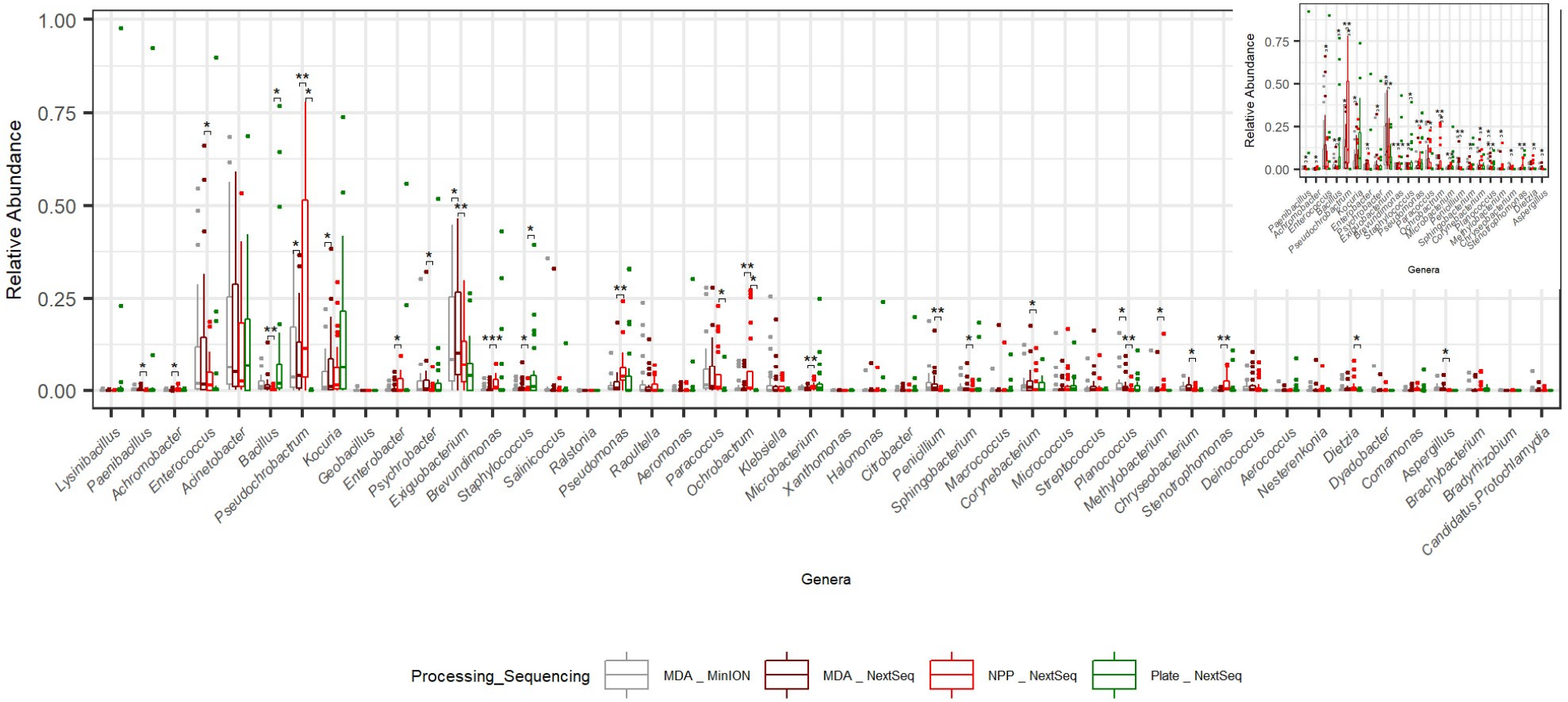
Venn diagram shows number of genera assigned per sequencing type, MinION > 5% relative abundance (M), NextSeq > 5% relative abundance (N), Sanger (S).



Supplemental Figure 5.6. Distribution of species level classification in environmental samples

[Supplemental Figure 5.6 continued]

Distribution of species classification by sequencer and pre-processing type. Significant differences highlighted. Controls not included in these, only environmental samples.



Supplemental Figure 5.7. Distribution of genera level classification in environmental samples.

[Supplemental Figure 5.7 continued]

Distribution of genera classification by sequencer and pre-processing type. Significant differences highlighted. Controls not included in these, only environmental samples.

Supplemental Table 5.1. Alignment of sequenced mock metagenome assembled genomes (QRY) to reference genomes from NCBI RefSeq (REF), using MUMmer.

	[REF]	[QRY]
[Sequences]		
TotalSeqs	10	104
AlignedSeqs	9(90.00%)	104(100.00%)
UnalignedSeqs	1(10.00%)	0(0.00%)
[Bases]		
TotalBases	19432784	17433642
AlignedBases	19095736(98.27%)	17362744(99.59%)
UnalignedBases	337048(1.73%)	70898(0.41%)
[Alignments]		
1-to-1	243	243
TotalLength	17330360	17043215
AvgLength	71318.35	70136.69
AvgIdentity	97.94	97.94

Supplemental Table 5.2. Counts of culturable bacteria from environmental samples, and their classification.

Source	Month	Culture type	CFU / Sample	CFU / Swab	Genera
Door	October	TBC	2.00E+04	4.00E+03	*Acinetobacter, Aerococcus, Aerococcus, Kocuria, Kocuria, Psychrobacter*
		ST	0.00E+00	0.00E+00	N/A
		SM	0.00E+00	0.00E+00	N/A
Table	October	TBC	3.34E+05	6.69E+04	*Aerococcus, Bacillus, Kocuria*
		ST	7.00E+01	1.40E+01	*Bacillus*
		SM	1.05E+04	2.10E+03	*Bacillus, Bacillus, Bacillus, Bacillus, Paenibacillus, Paenibacillus*
External Evaporator	October	TBC	5.88E+03	1.18E+03	*Bacillus, Lysinibacillus, Micrococcus*
		ST	3.50E+02	7.00E+01	*Bacillus*
		SM	2.94E+03	5.88E+02	*Bacillus, Bacillus, Bacillus*
External Dryer Bal Tank	October	TBC	9.38E+03	1.88E+03	*Bacillus, Bacillus, Bacillus, Chryseomicrobium, Exiguobacterium, Lysinibacillus*
		ST	7.00E+01	1.40E+01	N/A
		SM	7.00E+01	1.40E+01	N/A
Internal Dryer Bal Tank	October	TBC	3.22E+03	6.44E+02	*Bacillus, Lysinibacillus, Staphylococcus, Staphylococcus*
		ST	0.00E+00	0.00E+00	N/A
		SM	0.00E+00	0.00E+00	N/A
Wall	October	TBC	9.87E+03	1.97E+03	*Bacillus, Bacillus, Microbacterium, Rothia, Rothia*
		ST	7.70E+02	1.54E+02	*Bacillus*
		SM	1.82E+03	3.64E+02	*Oceanobacillus*
Gasket Flow Plate Seals	October	TBC	3.86E+08	7.71E+07	*Acinetobacter*
		ST	0.00E+00	0.00E+00	N/A
		SM	7.06E+04	1.41E+04	*Bacillus, Lysinibacillus*
Evaporator Drain	October	TBC	2.85E+08	5.70E+07	*Corynebacterium, Kocuria, Psychrobacter*
		ST	5.25E+02	1.05E+02	*Bacillus*
		SM	7.50E+04	1.50E+04	N/A
Neg ctrl	October	TBC	7.00E+01	1.40E+01	*Staphylococcus*
		ST	0.00E+00	0.00E+00	N/A
Swab		SM	0.00E+00	0.00E+00	N/A
Door	November	TBC	1.90E+04	3.81E+03	*Aerococcus, Bacillus, Kocuria, Staphylococcus, Staphylococcus, Staphylococcus, Staphylococcus, Staphylococcus*
		ST	3.15E+02	6.30E+01	*Bacillus*
		SM	3.15E+02	6.30E+01	*Bacillus*
Table	November	TBC	1.53E+05	3.06E+04	*Bacillus, Bacillus, Exiguobacterium, Exiguobacterium, Kocuria, Microbacterium*

		ST	2.80E+02	5.60E+01	N/A
		SM	1.05E+03	2.10E+02	*Bacillus, Lysinibacillus, Rothia*
External Evaporator	November	TBC	1.24E+07	2.48E+06	*Acinetobacter, Kocuria, Microbacterium, Microbacterium, Sphingobacterium*
		ST	7.00E+01	1.40E+01	*Bacillus*
		SM	3.57E+03	7.14E+02	*Bacillus, Bacillus, Bacillus, Kocuria*
External Dryer Bal Tank	November	TBC	9.28E+04	1.86E+04	*Acinetobacter, Bacillus, Bacillus, Chryseomicrobium, Paracoccus*
		ST	7.00E+01	1.40E+01	*Bacillus*
		SM	2.00E+03	3.99E+02	*Bacillus, Bacillus, Bacillus*
Internal Dryer Bal Tank	November	TBC	1.73E+04	3.46E+03	*Enterococcus, Enterococcus, Lysinibacillus*
		ST	0.00E+00	0.00E+00	N/A
		SM	1.26E+03	2.52E+02	*Bacillus*
Wall	November	TBC	2.18E+05	4.37E+04	*Kocuria, Kocuria, Lysinibacillus, Planococcus, Psychrobacter, Rothia, Rothia*
		ST	4.20E+02	8.40E+01	*Bacillus*
		SM	7.98E+03	1.60E+03	*Bacillus, Bacillus, Bacillus*
Gasket Flow Plate Seals	November	TBC	2.12E+07	4.24E+06	*Acinetobacter, Halomonas, Kocuria, Macrococcus*
		ST	7.00E+01	1.40E+01	*Bacillus*
		SM	5.85E+04	1.17E+04	*Bacillus, Lysinibacillus, Lysinibacillus*
Evaporator Drain	November	TBC	2.15E+09	4.30E+08	*Kocuria, Microbacterium, Planococcus*
		ST	1.47E+03	2.94E+02	*Bacillus, Bacillus*
		SM	3.53E+05	7.06E+04	*Bacillus, Bacillus, Corynebacterium*
Neg ctrl swab	November	TBC	2.10E+02	4.20E+01	*Bacillus, Kocuria, Rothia*
		ST	0.00E+00	0.00E+00	N/A
		SM	2.10E+02	4.20E+01	*Kocuria*
Door	December	TBC	1.61E+04	3.22E+03	*Arthrobacter, Micrococcus, Psychrobacillus, Psychrobacillus, Staphylococcus, Staphylococcus*
		ST	7.00E+01	1.40E+01	*Bacillus*
		SM	1.40E+02	2.80E+01	N/A
Table	December	TBC	2.10E+05	4.19E+04	*Bacillus, Kocuria, Rothia, Macrococcus, Rothia, Rothia, Staphylococcus*
		ST	6.30E+02	1.26E+02	*Bacillus*
		SM	8.40E+02	1.68E+02	N/A
External Evaporator	December	TBC	7.56E+03	1.51E+03	*Bacillus, Ornithinibacillus, Ornithinibacillus*
		ST	1.26E+03	2.52E+02	*Bacillus, Brevibacillus*
		SM	3.08E+03	6.16E+02	*Bacillus, Bacillus*

External Dryer Bal Tank	December	TBC	4.92E+04	9.84E+03	*Bacillus, Kocuria, Marinilactibacillus, Rothia, Salinicoccus*
		ST	3.15E+02	6.30E+01	*Bacillus*
		SM	3.50E+02	7.00E+01	N/A
Internal Dryer Bal Tank	December	TBC	7.00E+01	1.40E+01	*Lysinibacillus, Lysinibacillus*
		ST	0.00E+00	0.00E+00	N/A
		SM	7.00E+01	1.40E+01	*Bacillus*
Wall	December	TBC	2.17E+05	4.33E+04	*Aerococcus, Bacillus, Kocuria, Planococcus, Rothia*
		ST	1.05E+02	2.10E+01	*Thermoactinomyces*
		SM	1.47E+03	2.94E+02	*Oceanobacillus*
Gasket Flow Plate Seals	December	TBC	7.06E+07	1.41E+07	*Acinetobacter, Acinetobacter, Chryseobacterium*
		ST	8.40E+02	1.68E+02	*Bacillus*
		SM	4.57E+04	9.14E+03	*Bacillus, Bacillus, Exiguobacterium*
Evaporator Drain	December	TBC	2.94E+08	5.88E+07	*Aerococcus, Brevundimonas, Sphingobacterium*
		ST	8.72E+03	1.74E+03	*Brevibacillus*
		SM	1.17E+05	2.35E+04	*Bacillus, Bacillus*
Neg ctrl swab	December	TBC	1.40E+02	2.80E+01	*Bacillus*
		ST	0.00E+00	0.00E+00	N/A
		SM	7.00E+01	1.40E+01	*Bacillus*

Counts and classification of BHI cultured samples (TBC) and Mesophilic and thermophilic spore selected BHI cultured samples (SM and ST) following isolation of morphologically different isolates and identification by sanger sequencing of 16S region.

Supplemental Document 5.1

5.1.1 Taxonomic classification of negative controls

Species level taxonomic identification was performed on negative controls also. Many species identified were specific to negative controls with *Kribbia dieselivorans* and *Cytophagales bacterium* B6 being detected at a high relative abundance in MinION sequenced MDA negative controls, and *Paenibacillus fonticola* seen in high relative abundance in both MinION and Illumina sequenced MDA negative controls. There was also a high relative abundance of *Escherichia coli* in MDA negative controls, with *Salmonella enterica* in December samples, in both MinION sequences and corresponding Illumina sequences. *Ralsonia insidiosa* was also seen above 0.2% exclusively in negative controls. However, there was some overlap with some species identified in negative controls also identified in environmental samples. In particular, the swab negative control for both MDA MinION and MDA NextSeq from each month are similar to results generated from swabbing of the internal of the dryer balance tank, which are the environmental samples with the lowest environmental load (Supplemental Table 5.2). *Kocuria sp*., *Acinetobacter johnsonnii*, *Enterococcus casseliflavus*, *Klebsiella pneumoniae*, *Exiguobacterium sibricum*, *Enterococcus casseliflavus*, *Pseudochrobctrum sp* B5, *Enterobacter sp* HK169 and *Raoultella planticola* are all seen in negative controls (Figure 5.3).

There did appear to be some cross over between environmental sequences and negative controls particularly in environmental samples with low molecular loads. It must be considered that results for species classified in these samples could be false positives from cross over or contamination of sequences from other samples

at any stage of swabbing, extraction, amplification or sequencing. As this occurrence was noted in both MinION and NextSeq sequences it is unlikely to be due to barcode misassignment alone. There also was a possibility of false positive or incorrect classification for one or more species when multiple species of the same genera were classified and co-localised in the same samples. Another potential explanation could be index swopping.

5.1.2 Assignment of multiple species from single genera

Eleven genera have greater than 1 species assigned in MinION sequencing samples highlighted at > 5% relative abundance. These 11 genera contain 27 species out of a total of 59 species that are present in > 5% relative abundance in at least one sample. These species of the same genera are often co-localised in the same samples include 2 *Sphingobacterium*, 2 *Raoultella*, 2 *Nesterenkonia*, 2 *Micrococcus*, 3 *Kocuria*, 4 *Exiguobacterium*, 3 *Enterococcus*, 2 *Acinetobacter* species (Supplemental Figure 5.2, Figure 5.3). When Illumina sequences were included 24 genera had > 1 species assigned at > 5% relative abundance in at least one sample. These 24 genera contained 68 out of 108 species identified under these constraints (Figure 5.3).

Chapter 6. General Discussion

6.1 Discussion

The objective of this work was to apply novel sequencing-based methods to examine the microbiota of dairy processing facilities and products, with particular emphasis on spore-forming bacteria in powdered dairy products.

Chapter 1 evaluated common spore-forming contaminants of dairy powders, guidelines with respect to acceptable limits, advances in methods for detecting these spore-forming bacteria microorganisms, as well as advantages and limitations of these various methods.

The initial proof of concept study, described in Chapter 2, employed 16S rDNA amplicon sequencing of metagenomic DNA to investigate the impact of different ultra-high temperature (UHT) treatments in a pilot-scale plant on the dominant spore-forming populations in a batch of whey powder. Thermophilic sporeformers are prevalent in dairy powders (Watterson *et al.*, 2014) as well as powder processing lines (Cho *et al.*, 2018) and are regarded as indicators of poor hygiene (Burgess, Lindsay and Flint, 2010). UHT treatment can be employed to reduce thermophilic sporeformers and different thermal cycles can be employed in different processing runs. *Bacillus*, *Brevibacillus* and *Geobacillus* were identified as the dominant thermophilic genera in this powder. This result was similar to previous studies that identified *Bacillus* and *Geobacillus* as dominant genera in dairy powder (Miller *et al.*, 2015), while *Brevibacillus* has also been previously identified in dairy (Gopal *et al.*, 2015). Although fluctuations were evident, no significant differences relative abundances of genera due to UHT treatment were noted. However, significant differences in beta diversity between samples due to UHT

treatment were apparent. Decreased evenness of genera following ethidium bromide monoazide (EMA) treatment, to exclude DNA from dead cells from the analysis, significantly impacted some alpha diversity metrics. Overall, this study highlighted that UHT treatment of powder post-production differentially impacted thermophilic sporeformers present and demonstrated that high throughput DNA sequencing has the potential to be a tool to assess the composition of dairy products and provide insights into factors that influence the microbiology of food in processing.

Chapter 3 describes the use of shotgun metagenomic sequencing to investigate low-abundance mesophilic sporeformers in a powdered dairy product, whey powder, produced in one processing facility, monthly, over one year. Traditional culture-based detection and enumeration methods for spore-forming bacteria in dairy rely on a variety of temperature treatments, incubation temperatures and incubation conditions (Watterson *et al.*, 2014; Miller *et al.*, 2015). Different culture media including selective and chromogenic media to select for and identify potentially pathogenic bacteria (Weenk *et al.*, 1995; Tallent *et al.*, 2012) are also employed. The heavy reliance on culture-based assays has been highlighted as a concern in recent years. With a view to bridging the gap towards the application of sequencing technology to food quality and safety testing, shotgun metagenomic sequencing was used to determine the mesophilic spore content of 12 powdered dairy samples produced monthly, in the same processing facility, over one calendar year. For the purpose of comparison, one sample was examined in greater depth. This sample had its thermophilic spore content, as well as easily culturable

mesophilic and thermophilic spore content, examined. Three distinct mesophilic sporeformer population groups dominated the 12 powders. These groups were dominated by *Bacillus cereus*, *Bacillus licheniformis*/*Bacillus paralicheniformis* and a third more heterogenous group containing *Brevibacillus brevis*. Total thermophilic sporeformer taxonomic classification was different form the mesophilic taxonomic classification and the culturable thermophilic taxonomic classification, in the one sample analysed by all four approaches. Strain level analysis and functional gene potential resulted in the conclusion that low level transient contamination by potentially pathogenic sporeformers was occurring in these powders. There was no evidence for persistent contamination in the processing facility. Due to the sensitivity of these methods, new guidelines and standards may need to be introduced to ensure to ensure risks associated with detection of potential pathogens are adequately reflected in the interpretation of these results.

In Chapter 4, high throughput sequencing is used to track microorganisms from farms to a finished dairy powdered product. The impact of environmental factors, raw ingredients and processing environments on the microbiota of dairy products has previously been studied (Doyle *et al.*, 2017; Cho *et al.*, 2018; Fretin *et al.*, 2018; Wu *et al.*, 2018; Yeluri Jonnala *et al.*, 2018). However, more information is needed to develop a greater understanding of the factors that dictate changes in the dairy microbiota. In this study, 16S rDNA amplicon sequencing of metagenomic DNA was applied to investigate the microbial changes through a skim milk powder manufacturing process. Fresh samples from farm bulk tanks, collection tankers, milk silos, skimmed milk silos, a cream silo and powder samples were collected during

both mid and late lactation. Shotgun metagenomic sequencing was then applied to a subset of these samples to allow further investigation. Raw milk on farms and in collections tankers had a very diverse microbiota. However, upon pooling and storage in a whole milk silo at the processing facility, *Pseudomonas fluorescens* and *Acinetobacter baumannii* became dominant. This is consistent with a previous study that noted that large scale silo storage led to the convergence of the microbiota to one of two community types (Kable *et al.*, 2016). Proliferation of low levels of thermophilic bacteria present either in the raw ingredients or the processing facility, namely *Thermus thermophilus* and *Geobacillus sp*., on one of the processing days, resulted in their dominance in the corresponding final powder. As noted above, thermophilic bacteria are capable of withstanding high temperature processing and are a concern for the dairy industry (Miller *et al.*, 2015; Cho *et al.*, 2018). Overall this study provides an insight into the changes in the microbiota of dairy through a powder manufacturing process, on distinct sampling days. With routine implementation of these methods, a greater understanding of the factors that dictate change in the dairy microbiota and lead to different species being dominant in a final product can be determined. This in turn could lead to informed decisions regarding product fate to reduce further reduce risk, increase safety and reduce economic losses.

Chapter 5 describes an investigation involving the use of a newer, portable, long read DNA sequencing platform, i.e. Oxford Nanopore MinION, to monitor food processing environments, and compares the outputs relative to those provided by Illumina-based sequencing, and culture-based techniques. Although long read

portable sequencing has been used for rapid detection of pathogens in clinical settings (Charalampous *et al.*, 2019), it has yet to be used in a food processing facility. Initially the ability of MinION-based approaches to distinguish between the components a very simple mock metagenomic community of related food-associated spore-forming bacteria was assessed. Multiple displacement amplification (MDA) was then applied to metagenomic DNA extracted from environmental swabs from a processing facility. These samples were then sequenced using the rapid sequencing kits designed for use with the MinION sequencer as well as Illumina NextSeq sequencer. DNA that was not subjected to pre-processing/MDA was also sequenced using Illumina-based sequencing. In addition, culture-based analysis was performed, colonies were washed from the resultant plates and metagenomic DNA from these easily culturable microorganisms was extracted and sequenced using the Illumina NextSeq. Finally, Sanger sequencing of a region of the 16S genera was performed on easily culturable isolates. All resulted were compared. Overall taxonomic classification from Oxford Nanopore sequences was comparable with corresponding Illumina sequences. MDA did introduce some bias, as did culture-based analysis. Metagenome assembled genomes (MAGs) were also recovered from a hybrid assembly of MDA MinION and Illumina reads. Three of the seven high quality MAGs appeared to be previously uncharacterised species. These MAGs will add to the knowledge of food processing contaminants and accelerate classification in the future. More importantly, the relevance of portable MinION sequencing to food processing facilities was established.

6.2 Conclusion

Overall this thesis aimed to investigate the potential to apply novel sequencing-based methods to examine the microbiota of dairy processing, with particular emphasis on spore-forming bacteria in powdered dairy products. The methods used in this thesis can be used directly, or provide the foundation for future advances, to provide a more in depth understanding of the dairy microbiota and, in turn, make timely, informed decisions to maximise food safety and economic output.

6.3 References

Burgess, S. A., Lindsay, D. and Flint, S. H. (2010) 'Thermophilic bacilli and their importance in dairy processing', *Int J Food Microbiol,* 144(2), pp. 215-25. DOI: 10.1016/j.ijfoodmicro.2010.09.027

Charalampous, T., Kay, G. L., Richardson, H., Aydin, A., Baldan, R., Jeanes, C., Rae, D., Grundy, S., Turner, D. J., Wain, J., Leggett, R. M., Livermore, D. M. and O'Grady, J. (2019) 'Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection', *Nat Biotechnol*. DOI: 10.1038/s41587-019-0156-5

Cho, T. J., Kim, H. W., Kim, N. H., Park, S. M., Kwon, J. I., Kim, Y. J., Lee, K. W. and Rhee, M. S. (2018) 'New insights into the thermophilic spore-formers in powdered infant formula: Implications of changes in microbial composition during manufacture', *Food Control,* 92, pp. 464-470. DOI: 10.1016/j.foodcont.2018.05.036

Doyle, C. J., Gleeson, D., O'Toole, P. W. and Cotter, P. D. (2017) 'Impacts of seasonal housing and teat preparation on raw milk microbiota: a high-throughput sequencing study', *Appl Environ Microbiol,* 83(2). DOI: 10.1128/AEM.02694-16

Fretin, M., Martin, B., Rifa, E., Isabelle, V. M., Pomies, D., Ferlay, A., Montel, M. C. and Delbes, C. (2018) 'Bacterial community assembly from cow teat skin to ripened cheeses is influenced by grazing systems', *Sci Rep,* 8(1), pp. 200. DOI: 10.1038/s41598-017-18447-y

Gopal, N., Hill, C., Ross, P. R., Beresford, T. P., Fenelon, M. A. and Cotter, P. D. (2015) 'The prevalence and control of Bacillus and related spore-forming bacteria in the dairy industry', *Front Microbiol,* 6, pp. 1418. DOI: 10.3389/fmicb.2015.01418

Kable, M. E., Srisengfa, Y., Laird, M., Zaragoza, J., McLeod, J., Heidenreich, J. and Marco, M. L. (2016) 'The Core and Seasonal Microbiota of Raw Bovine Milk in Tanker Trucks and the Impact of Transfer to a Milk Processing Facility', *MBio,* 7(4). DOI: 10.1128/mBio.00836-16

Miller, R. A., Kent, D. J., Watterson, M. J., Boor, K. J., Martin, N. H. and Wiedmann, M. (2015) 'Spore populations among bulk tank raw milk and dairy powders are significantly different', *J Dairy Sci,* 98(12), pp. 8492-504. DOI: 10.3168/jds.2015-9943

Quigley, L., O'Sullivan, D. J., Daly, D., O'Sullivan, O., Burdikova, Z., Vana, R., Beresford, T. P., Ross, R. P., Fitzgerald, G. F., McSweeney, P. L., Giblin, L., Sheehan, J. J. and Cotter, P. D. (2016) 'Thermus and the Pink Discoloration Defect in Cheese', *mSystems,* 1(3). DOI: 10.1128/mSystems.00023-16

Tallent, S. M., Kotewicz, K. M., Strain, E. A. and Bennett, R. W. (2012) 'Efficient isolation and identification of Bacillus cereus group', *J AOAC Int,* 95(2), pp. 446-51. DOI: 10.5740/jaoacint.11-251

Watterson, M. J., Kent, D. J., Boor, K. J., Wiedmann, M. and Martin, N. H. (2014) 'Evaluation of dairy powder products implicates thermophilic sporeformers as the primary organisms of interest', *J Dairy Sci,* 97(4), pp. 2487-97. DOI: 10.3168/jds.2013-7363

Weenk, G. H., van den Brink, J. A., Struijk, C. B. and Mossel, D. A. A. (1995) 'Modified methods for the enumeration of spores of mesophilic Clostridium species in dried foods', *International Journal of Food Microbiology,* 27(2), pp. 185-200. DOI: 10.1016/0168-1605(94)00164-2

Wu, S., Jiang, Y., Lou, B., Feng, J., Zhou, Y., Guo, L., Forsythe, S. J. and Man, C. (2018) 'Microbial community structure and distribution in the air of a powdered infant formula factory based on cultivation and high-throughput sequence methods', *J Dairy Sci*, 101(8), pp. 6915-6926. DOI: 10.3168/jds.2017-13968

Yeluri Jonnala, B. R., McSweeney, P. L. H., Sheehan, J. J. and Cotter, P. D. (2018) 'Sequencing of the Cheese Microbiome and Its Relevance to Industry', *Front Microbiol,* 9, pp. 1020. DOI: 10.3389/fmicb.2018.01020

Acknowledgements

I would like to thank each and every person who has helped me over the last number of years. Without the constant support, help and guidance of those around me this thesis would not have been possible.

Sincere gratitude to my supervisory team for allowing me to undertake this research. First and foremost, I would like to thank my primary supervisor Prof. Paul Cotter. Words cannot do justice the gratitude I hold for all the time and effort he has invested in me throughout the last number of years. For his continued guidance, patience and generosity with time I will forever be grateful. His optimism, enthusiasm, generosity and diplomacy while juggling so many tasks were beyond inspirational. I would like to thank Dr Conor Feehily for his constant support, guidance and advice throughout the last number of years. His wealth of knowledge and ability to spot a typo were impressive. I would like to thank my university supervisor Prof. Colin Hill. It was a great reassurance knowing he was only an email away; he was always helpful and supportive throughout the years. Thank you all sincerely.

I would like to thank the Department of Agriculture, Food and the Marine for providing the funding for this project. I would like to thank Dr John Tobin and everyone involved in the SACCP project, for help sourcing samples and organising project meetings. I would like to thank all the companies that donated samples to the various studies or allowed me into their facilities. Without their generosity, this thesis would not have been possible. I would like to thank Dr David Gleeson and Lizandra Paludetti for sourcing and organising delivery of samples (Chapter 4). I

would also like to thank Fang Li for providing and helping process samples (Chapter 2).

To everyone at Teagasc that has made my time there enjoyable and educational, especially all members of Vision 1, past and present, and those that resided in the student office. Sincere thanks for the support and guidance shown by all. Sincere thanks to Dr Fiona Crispie and Laura Finnegan for sequencing the Illumina runs, as well as bestowing a small amount of their infinite sequencing knowledge onto me. Likewise, Dr Orla O'Sullivan sincere thanks for helping me with bioinformatic pipelines as well as for encouraging me to present my results. Sincere thanks to Elaine Lawton for always being there to ask questions and giving guidance in the lab. Drs Calum Walsh, Wiley Barton and Aaron Walsh, sincere thanks for helping me with bioinformatics as well as having chats in the office and over tea. Min, sincere thanks for helping me over the final year of my PhD, both swabbing and preparing samples for analysis. John, thanks for always being helpful, and often providing comic relief at tea break. Chloe, thanks for all the laughs, and for always trying to bring people together, opening your house and organising events. Cathy, thanks for all the chats, the walks, trying in vain to teach me to be cool and always listening. Ciara, thank you for being an inspiration in your work ethic and outreach activities. Thank you for your help and support from day one, but particularly in our final year when we were both finishing up. Thanks to the three of you for the chats, coffee, tea, cake and company at lunch breaks. You all played a part in making it a very enjoyable few years.

Special thanks to my parents, who were happily enjoying an empty nest, until I came bounding back in and set up camp. The patience and support they have shown me throughout these years, and throughout my life has been endless. Their support despite their bewilderment at my desire for continuous learning as well as their never-ending doubt that I will ever get a 'real job', has all cumulatively shaped the person I am today, and helped me finish this thesis. Thank you sincerely. A special mention must be given to the dogs Ali and Ted, a walk with either of them, or both if I was feeling brave, took my mind off the PhD, and was a great escape.

Finally Niall, heartfelt thank you for your continuous support. Your patience and commitment to the long commutes for visits, or warm welcome on my arrival made the end of a tough week all the better. Thank you sincerely for unceasing encouragement, particularly in the final few months.

To anyone that has helped me in any way; thanks a million!